Compressed Sensing for Chemistry

Citation

Share Your Story

Compressed Sensing for Chemistry

A dissertation presented
by
Jacob Nathan Sanders
to
The Committee in Chemical Physics

in partial fulfillment of the requirements
for the degree of
Doctor of Philosophy
in the subject of
Chemical Physics

Harvard University
Cambridge, Massachusetts

April 2016

Dissertation Advisor   Author

Alán Aspuru-Guzik   **Jacob Nathan Sanders**

Compressed Sensing for Chemistry

Abstract

Many chemical applications, from spectroscopy to quantum chemistry, involve measuring or computing a large amount of data, and then compressing this data to retain the most chemically-relevant information. In contrast, compressed sensing is an emergent technique that makes it possible to measure or compute an amount of data that is roughly proportional to its information content. In particular, compressed sensing enables the recovery of a sparse quantity of information from significantly undersampled data by solving an ℓ_1-optimization problem. This thesis represents the application of compressed sensing to problems in chemistry.

The first half of this thesis is about spectroscopy. Compressed sensing is used to accelerate the computation of vibrational and electronic spectra from real-time time-dependent density functional theory simulations. Using compressed sensing as a drop-in replacement for the discrete Fourier transform, well-resolved frequency spectra are obtained at one-fifth the typical simulation time and computational cost. The technique is generalized to multiple dimensions and applied to two-dimensional absorption spectroscopy using experimental data collected on atomic rubidium vapor. Finally, a related technique known as super-resolution is applied to open quantum systems to obtain realistic models of a protein environment, in the form of atomistic spectral densities, at lower computational cost.

The second half of this thesis deals with matrices in quantum chemistry. It presents a new use of compressed sensing for more efficient matrix recovery whenever the calculation of individual matrix elements is the computational bottleneck. The technique is applied to the

computation of the second-derivative Hessian matrices in electronic structure calculations to obtain the vibrational modes and frequencies of molecules. When applied to anthracene, this technique results in a threefold speed-up, with greater speed-ups possible for larger molecules. The implementation of the method in the Q-Chem commercial software package is described. Moreover, the method provides a general framework for bootstrapping cheap low-accuracy calculations in order to reduce the required number of expensive high-accuracy calculations.

Contents

vi

Citations to previously published work

Apart from minor modifications, Chapters 2–5 appeared as the following publications [13, 155, 115, 154]:

"Application of Compressed Sensing to the Simulation of Atomic Systems." Xavier Andrade, Jacob N. Sanders, and Alán Aspuru-Guzik. *Proceedings of the National Academy of Sciences USA*. Volume 109, Number 35: 13928–13933. 2012.

"Compressed Sensing for Multidimensional Spectroscopy Experiments." Jacob N. Sanders, Semion K. Saikin, Sarah Mostame, Xavier Andrade, Julia R. Widom, Andrew H. Marcus, and Alán Aspuru-Guzik. *The Journal of Physical Chemistry Letters*. Volume 3, Number 18: 2697–2702. 2012.

"More Accurate and Efficient Bath Spectral Densities from Super-Resolution". Thomas Markovich, Samuel M. Blau, John Parkhill, Christoph Kreisbeck, Jacob N. Sanders, Xavier Andrade, and Alán Aspuru-Guzik. arXiv preprint 1307.4407. 2013.

"Compressed Sensing for the Fast Computation of Matrices: Application to Molecular Vibrations." Jacob N. Sanders, Xavier Andrade, and Alán Aspuru-Guzik. *ACS Central Science*. Volume 1, Number 1: 24–32. 2015.

I also contributed to one other publication during the course of my doctorate which did not make it into this dissertation [116]:

"Benchmarking Compressed Sensing, Super-Resolution, and Filter Diagonalization." Thomas Markovich, Samuel M. Blau, Jacob N. Sanders, and Alán Aspuru-Guzik. *International Journal of Quantum Chemistry*. In Press. 2016. arXiv preprint 1502.06579.

Acknowledgments

First and foremost, I want to thank my adviser, my teacher, and my friend, Alán Aspuru-Guzik. While Harvard has many great scientists, Alán is also a wonderful human being. No one else I know is so talented at motivating people, at appealing to their better natures, and at making them excited and proud to produce the highest quality work possible. When things are rough, he lends a sympathetic ear; when things are good, he is your biggest cheerleader; when things are slow, he offers a friendly but firm push to make progress; and when you get a result, he lights up with excitement. He has a unique understanding of what makes people "tick," and he knows how to fine-tune his mentorship to match the way you are feeling at any given moment. On a personal note, I have cherished our morning runs through Cambridge and Somerville. The chance to talk almost every day, as two friends sharing their lives in pursuit of the truth, encapsulates for me the ancient Greek concept of the good life.

I am particularly indebted to Xavier Andrade for "showing me the ropes" during the first half of my Ph.D. As I co-authored my first papers under his expert guidance, I always knew I was in good hands. More than anyone else, he taught me how to write like a scientist, and I am still amazed by how much precise meaning he could pack into a single sentence without overwhelming the reader. I miss his palpable excitement for all things computational, the wizard-like tricks he could pull off at the Linux command line, and most of all his generous mentorship. My debt to Xavier is one I can never hope to repay.

I am also grateful to Joel Yuen-Zhou for taking me off the (metaphorical) street and mentoring me during my earliest days in the Aspuru-Guzik group. Joel made quantum mechanics feel real and alive, and I remember my excitement at finishing my first "grown-

up" perturbation theory calculations under his guidance. Leading by example, Joel taught me to think rigorously, strip away needless complexity, and get at the heart of hard problems.

The Aspuru-Guzik group would be unrecognizable to me without the day-to-day humor and friendship of Joey Goodknight. My officemate of more than three years, Joey made the daily grind of a Ph.D. worth it. I could joke, share stories, and—on many occasions— exchange knowing glances. Even when Joey's jokes are at your expense, they are inclusive, friendly, and warm. I will always remember how much fun we had together on our trip to Paris with the long layover in Vienna. The group would also be a completely different place without Thomas Markovich, who is an incredible collaborator, conversationalist, and friend. Thomas's strengths perfectly complement my weaknesses and, as a result, I could always count on him to help me when I needed help the most.

Within the Aspuru-Guzik group, I have benefited immensely from the electronic structure expertise of Dmitrij Rappoport, the complaints club of Steven Lopez, the dry Russian wit of Semion Saikin, the masterful teaching of Jacob Krich and Jarrod McClean, the advice of Leslie Vogt, the friendship of Carolina Román, and the dry British wit, career advice, and general support of Peter Love. I have enjoyed sharing an office with Stéphanie Valleau and Sam Blau, and getting to run with Nico Sawaya and David Gelbwaser. Other group members whom I have been fortunate to learn from include Ryan Babbush, Martin Blood-Forsythe, Adrian Jinich, Ivan Kassal, Ian Kivlichan, Roberto Olivares, Alejandro Perdomo-Ortiz, Patrick Rebentrost, Jonathan Romero Fontalvo, Benjamin Sanchez, Sangwoo Shim, Teresa Tamayo, Jennifer Wei, Jon Welch, James Whitfield, Doran Bennett, Sergio Boixo, Alex Eisfeld, Johannes Hachmann, Christoph Kreisbeck, Sarah Mostame, John Parkhill, Borja Peropadre, Ed Pyzer-Knapp, César Rodríguez-Rosario, David Tempel, Mark Watson, Man Hong Yung, Dmitry Zubarev, Carlos Amador-Bedolla, Ian Dunn,

Dipti Jasrasaria, Trevor Lutzow, Anna Shin, and Marlon Cummings.

Beyond the Aspuru-Guzik group, I am grateful to Yihan Shao for hosting me for a stimulating scientific visit to the NIH during which he generously helped me to implement part of my Ph.D. work into the Q-Chem electronic structure software package. I am similarly grateful to Angel Rubio for hosting a winter school in Benasque on time-dependent density functional theory, which taught me all of the background I needed to co-author my first paper, and later for welcoming me during a scientific visit to the European Center for Theoretical Spectroscopy in San Sebastian. I am also grateful to Andy Marcus for a fruitful collaboration during the early part of my Ph.D. I appreciate the generosity of Adam Cohen and Yue Lu in taking the time to review this dissertation, serving on my dissertation committee, and offering me valuable feedback and new ideas for my research every year.

Prior to and during my Ph.D., I have been fortunate to serve as a Head Teaching Fellow in various courses with Andrew Myers, Tobias Ritter, Adam Cohen, Logan McCarty, and Hongkun Park. I have learned so many tricks, both large and small, from all of them. Working closely with so many outstanding teachers has inspired my own teaching, and I often find myself trying to channel the most memorable parts of their lectures when I explain concepts to students. The advice and encouragement of Logan McCarty is one of the principal reasons why I decided to pursue a Ph.D. in the first place. While I never managed to teach directly with Gregg Tucci, Gregg has been an incredible supporter and soundboard for all things related to teaching, and every single time I have walked into his office I have walked out happier. I am also deeply indebted to the hundreds of students I have taught over the years. Their questions forced me to understand how disparate subfields of chemistry fit together, and led me to explore areas I never would have had the chance to think about otherwise. Many of the happiest hours of my Ph.D. were spent in the classroom,

where I always felt like I was surrounded by friends excited to explore chemistry with me. I am especially happy to have remained good friends with László Seress, who I met on the first day of his freshman year at Harvard when he boldly enrolled in a sophomore class, and who I now have the pleasure of watching as he begins his own graduate career in theoretical chemistry.

For the last three years, I had the opportunity to take the most incredible two-week "vacation" each June: I got to serve as an instructor at the Chemistry Study Camp, held at the U.S. Air Force Academy in Colorado Springs, which brings together the top 20 high-school chemistry students across the country to prepare them for the International Chemistry Olympiad competition. I am grateful to Kelli Slunt, Michael Danahy, Melissa Barranger-Mathys, Amiee Modic, Patrick Chan, Cecilia Hernandez, and Kim Gardner for being outstanding co-mentors and coordinators. Working with brilliant young students who are crazy-excited about chemistry was a thrilling experience I cannot adequately describe in words. It puts the entire academic enterprise into perspective and reminds me what it is all *for*, and nothing else makes me more optimistic about the future of chemistry. I can't help but smile when I hear about materials science from Derek Wang, plunge into ligand field theory with Saaket Agrawal, discuss an off-beat chemistry paper with Alvin Hsu, or learn of the latest crazy thing David Wang managed to do with copper. Among the students, I have never met anyone else with the intelligence, intensity, and persistence of Stephen Ting, who has become one of my closest friends, and with whom I can talk about anything ranging from organic chemistry to thermodynamics at all hours of the day or night. No other student has taught me as much, or motivated me as much, as he has.

As for my own education, I was lucky to attend a high school with what must be one of the strongest chemistry departments in the country. It was in the classrooms of David

Ostfeld, Gull-Maj Roberts, Todd Crane, and Laura Crane that I fell in love with chemistry and knew that I wanted to become a chemist. The impact on students of having four inspired chemistry teachers, year after year, cannot be overstated. Going even further back, it was in the elementary school classroom of Sandra McGarry that I first enjoyed science and learned of the periodic table.

No one makes it through a Ph.D. without the rock-solid support of a large network of friends, people who you can just hang out with without any conditions. Though there are too many to name, all of them provided bright spots on difficult days without even knowing it. Whereas most of my friends had some housing drama at some point during their Ph.D., I have been lucky to live with two of the best housemates throughout my entire graduate career, Brian Lee and John Janetzko. Brian taught me how to optimize my life and optimize my computer. He is also friendly, off-beat, and just *different* enough to be really interesting. He is a careful thinker with refreshingly new ideas. In both academic and athletic realms, John is the most inspiring person I know: he could motivate a *rock* to adopt an active lifestyle while earning a Ph.D. Luckily, I am not a rock, but you get the idea. Among all of my friends, no one has been by my side longer than Anna Shneidman. She has stuck with me through middle school, high school, college, and graduate school. For 16 years, we have gone on long walks where we have talked about everything and, just like 16 years ago, we still begin each walk by asking each other, "How's life?" It will not be easy for me to move across the country and separate from all of these people.

Last but not least, I am forever grateful for the unconditional love of my family. My grandparents, Jack Sanders, Eva Sanders, Carrie Ignatin, and Irving Ignatin, overcame so many obstacles so that their children and grandchildren could enjoy a free and fulfilling life. I am so glad I have grown close to my sister, Rebecca Sanders, and she is the one person in

For my parents
For my sister
For my students

Chapter 1

Introduction

A spectrum is measured but one is interested in just the peaks. An expensive matrix is computed but one cares only about obtaining accurate eigenvalues. A large microscope field-of-view is acquired but one is interested only in a few active locations. An entire magnetic resonance image is constructed bit-by-bit, but the medically-relevant information is in the locations where the density of the nuclei changes.

All of these scenarios demonstrate the same paradigm which dominates chemistry: we measure or compute a large amount of data, and then compress this data to focus on the chemically-relevant information. One might imagine an alternative approach in which one directly measures or computes the chemical information. The potentially fatal problem, of course, is that typically one does not know *a priori* where the information is to be found within the large set of data or even how much information is in the data.

Nevertheless, an emergent technique in applied mathematics known as compressed sensing has shown that this problem is far from fatal, and that it is possible to measure or compute an amount of data that is *proportional* to the amount of information present in the data [34, 57, 35, 172]. This thesis represents the application of this new compressed sensing paradigm to problems in chemistry. In a variety of chemical applications, this thesis

suggests more efficient approaches to acquiring chemical information by ensuring that the amount of data to be calculated or collected is proportional to the information content.

While this thesis focuses on applications to spectroscopy and electronic structure, compressed sensing—which has been termed a paradigm shift from "measure what *can* be measured" to "measure what *should* be measured" [172]—reaches far beyond chemistry, with an enormous number of other applications springing up across science and engineering. Experimental applications of compressed sensing include super-resolution microscopy [205], neuronal circuitry [82], quantum state and process tomography [69, 159, 63, 15], and a carbon nanotube-based infrared camera [169]. On the computational side, compressed sensing is becoming a tool for constructing physics models [128] and for modeling sparse dynamics of partial differential equations [157]. Such a list of applications could be reconstructed many times over, using different applications each time. We begin with a brief mathematical introduction to compressed sensing.

1.1   Compressed Sensing

Consider a column-vector $\mathbf{x}$ containing n unknown variables, and suppose we measure linear combinations of the unknown variables and collect them in the column-vector $\mathbf{b}$ containing m such measurements. We can formulate the problem of solving for the unknowns as a linear inversion problem,

$$A\mathbf{x} = \mathbf{b}, \tag{1.1}$$

where A is an $m \times n$ matrix of coefficients relating the unknown vector $\mathbf{x}$ to the measurement vector $\mathbf{b}$. From standard linear algebra, we know that if $\mathbf{x}$ contains n unknowns, then we need $\mathbf{b}$ to contain $m = n$ linearly independent measurements at which point equation (1.1)

becomes a square system which admits the unique solution

$$\mathbf{x} = A^{-1}\mathbf{b}. \tag{1.2}$$

When the number of linearly-independent measurements m is less than the number of unknowns n, equation (1.1) becomes an underdetermined system of equations admitting an infinite number of solutions. Attempting to solve for n unknowns with fewer than n measurements is, in the general case, a hopeless endeavor. The key insight of compressed sensing is that in the special case when the unknown vector $\mathbf{x}$ is *sparse*, meaning it contains only a small number of nonzero values, it becomes possible to solve for $\mathbf{x}$ with fewer than n measurements.

In compressed sensing, one simply finds the *sparsest* vector $\mathbf{x}$ which satisfies the underdetermined system in equation (1.1). In other words, one solves the following ℓ_0 optimization problem, [34, 57, 35, 16]

$$\min \|\mathbf{x}\|_0 \ \text{ subject to } A\mathbf{x} = \mathbf{b}, \tag{1.3}$$

where

$$\|\mathbf{x}\|_0 = \sum_{i=1}^{n} |x_i|^0 \tag{1.4}$$

is the zero-norm which simply counts the number of non-zero elements in the vector $\mathbf{x}$.

Consider what happens to the solution to the optimization problem (1.3) when the unknown vector $\mathbf{x}$ is known to be sparse. When the measurement vector $\mathbf{b}$ contains very few measurements, so that the system of equations $A\mathbf{x} = \mathbf{b}$ is highly underdetermined, it is likely that the optimization problem (1.3) admits a solution for $\mathbf{x}$ that is *sparser* than the true solution. However, as more measurements are collected, so that the vector $\mathbf{b}$ grows in size, eventually the solution to the optimization problem (1.3) will converge to the correct

sparse solution for $\mathbf{x}$. Moreover, if $\mathbf{x}$ is sparse enough, this convergence can happen long there are as many measurements as unknowns, in other words, long before $A\mathbf{x} = \mathbf{b}$ is a square system admitting a unique solution. In fact, the number of required measurements m scales linearly with the number of non-zeros in the unknown vector $\mathbf{x}$, also known as the *sparsity* of $\mathbf{x}$, as discussed further near the end of this section.

Unfortunately, in practice, solving the optimization problem (1.3) turns out to be NP-hard [16], but luckily it has been proven to be equivalent to the following ℓ_1 optimization problem [34, 57, 35, 16]

$$\min ||\mathbf{x}||_1 \ \text{subject to } A\mathbf{x} = \mathbf{b}, \tag{1.5}$$

where

$$||\mathbf{x}||_1 = \sum_{i=1}^{n} |x_i|^1 \tag{1.6}$$

is simply the sum of the absolute values of the entries in $\mathbf{x}$. This ℓ_1 optimization problem, which can be formulated as a linear program, is what one actually solves when implementing compressed sensing.

To understand how compressed sensing works, and in particular why the ℓ_1 optimization problem in equation (1.5) is equivalent to the ℓ_0 optimization problem in equation (1.3), it is useful to examine a simple low-dimensional example [16]. Suppose the vector $\mathbf{x} = \left[\begin{smallmatrix} x_1 \\ x_2 \end{smallmatrix}\right]$ contains only two unknowns and is sparse, which in this simple example we take to mean that $x_2 = 0$. Figure 1.1 illustrates the two-dimensional space of possible solutions for $\mathbf{x}$ and represents the solution by a dot. The standard approach to solving the system $A\mathbf{x} = \mathbf{b}$ is to collect *two* linearly-independent measurements in $\mathbf{b}$ and then invert the square system of equations. Each measurement corresponds to a line and the point at which these two lines cross yields the solution for $\mathbf{x}$, as shown in Figure 1.2.
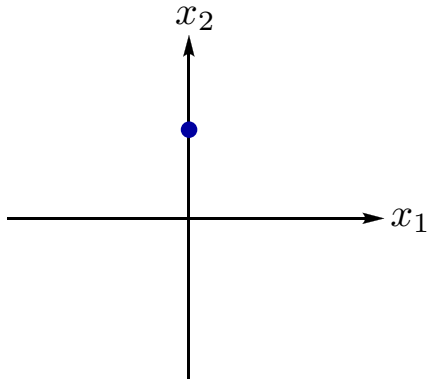
Figure 1.1: This figure shows a sparse vector $\mathbf{x}$ with $x_2 = 0$ to be recovered.
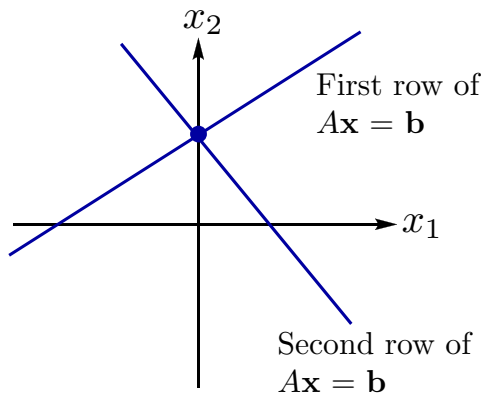
Figure 1.2: This figure illustrates the conventional recovery of sparse vector $\mathbf{x}$ by solving a square system of linearly-independent equations.
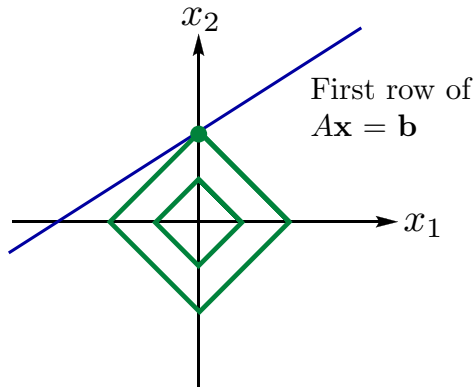
Figure 1.3: This figure illustrates the recovery of sparse vector $\mathbf{x}$ by applying compressed sensing via ℓ_1 optimization. Each diamond represents constant ℓ_1 norm, and the location where the smallest possible diamond intersects the line is the solution to the optimization problem in equation (1.5). We see that $\mathbf{x}$ is recovered correctly [16].

In compressed sensing, the idea is to make fewer measurements in $\mathbf{b}$ and then employ the ℓ_1 optimization in equation (1.5) to solve for $\mathbf{x}$. Suppose we make only one measurement, b_1, thus obtaining only one line in Figure 1.3. In principle, the solution for $\mathbf{x}$ could lie anywhere along this line. The diamonds in Figure 1.3 represent lines of constant ℓ_1 norm, so the smallest diamond to intersect the line represents the solution to the ℓ_1 optimization problem in equation (1.5) [16]. As this simple cartoon shows, compressed sensing via ℓ_1 optimization recovers the correct solution for $\mathbf{x}$ using only one measurement rather than two. In more realistic high-dimensional settings, the computational savings can be much greater.

Figure 1.4 shows that the ℓ_1 norm is crucial for solving the compressed sensing problem. If the ell_1 norm is replaced by the ℓ_2 sum-of-squares norm, which corresponds to replacing the diamonds with circles, the recovered solution for $\mathbf{x}$ is incorrect and no longer sparse [16]. Thus, the ℓ_1 norm represents a happy medium: it is "sharp" enough to favor sparse solutions
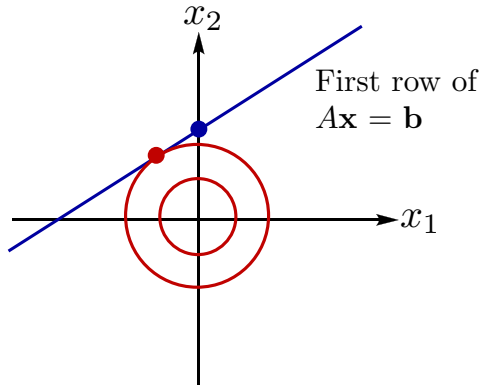
Figure 1.4: Replacing the ℓ_1 optimization by an ℓ_2 optimization precludes the correct recovery of sparse vector $\mathbf{x}$, as the smallest possible circle does not intersect the line in the correct location [16].

(which is not the case with norms larger than the 1-norm) and it is not NP-hard to minimize (which is not the case with norms smaller than the 1-norm).

We note that in solving the optimization problem in equation (1.5), the equality of the underdetermined system $A\mathbf{x} = \mathbf{b}$ is enforced to a specified tolerance, so the true optimization problem which is solved is [184]

$$\min ||\mathbf{x}||_1 \text{ subject to } ||A\mathbf{x} - \mathbf{b}||_2 < \eta, \tag{1.7}$$

for some small tolerance η usually between 10^{-4} and 10^{-8}.

If we regard $\mathbf{x}$ as a vector expressed in a basis $\{\phi_i\}$,

$$\mathbf{x} = \sum_i x_i \phi_i, \tag{1.8}$$

and $\mathbf{b}$ as a vector expressed in a basis $\{\psi_j\}$,

$$\mathbf{b} = \sum_j b_j \psi_j, \tag{1.9}$$

so that A becomes the change-of-basis matrix from the $\{\phi_i\}$ basis to the $\{\psi_j\}$ basis, then

we can define the *coherence* μ of the change-of-basis transformation as [58, 31, 35]

$$\mu = \sqrt{n} \max_{1 \leq i,j \leq n} \frac{\langle \phi_i | \psi_j \rangle}{\sqrt{\langle \phi_i | \phi_i \rangle \langle \psi_j | \psi_j \rangle}}, . \tag{1.10}$$

a number which ranges from a minimum of 1 when each vector in $\{\phi_i\}$ avoids as much as possible overlapping with any vector in $\{\psi_j\}$ to a maximum of $\sqrt{n}$ when there exist vectors in the $\{\phi_i\}$ and $\{\psi_j\}$ that completely coincide. In the former case, we say the two bases are *incoherent* while in the latter case, we say they are *coherent*. In terms of this coherence parameter, the number of measurements m in the vector $\mathbf{b}$ needed to accurately recover the unknown vector $\mathbf{x}$ is proven to scale as [31, 35]

$$m \sim \mu^2 s \log n, \tag{1.11}$$

where s is the *sparsity* of $\mathbf{x}$ (number of nonzeros) and μ is the *coherence* parameter. This formula encapsulates two features of compressed sensing. First, the number of measurements required to recover a sparse vector $\mathbf{x}$ scales directly with its sparsity. Second, the fewest measurements are required when the sparse basis $\{\phi_i\}$ and the measurement basis $\{\psi_j\}$ are as *incoherent* as possible. In more informal language, the optimal situation for compressed sensing is when A "scrambles up" all the elements of $\mathbf{x}$ to give $\mathbf{b}$, as this ensures that each measured element of $\mathbf{b}$ is maximizing the amount of information provided about *all* the elements of $\mathbf{x}$.

1.2   Application to Image Processing

One of the most important applications of compressed sensing is to the rapid acquisition of images, where the technique enables a dramatic reduction in the number of measurements required to obtain a well-resolved image [35, 16]. In this section, we use image processing to provide a visual illustration of how compressed sensing works in practice.

Images are composed of pixels and most of these pixels are nonzero; as such, images are not immediately sparse. However, most pixels in an image are quite similar to their neighbors; that is, most areas of an image do not involve pixels undergoing drastic changes on very short length scales. As a result, the two-dimensional discrete cosine transform (a real analog of the complex discrete Fourier transform) of an image is typically quite sparse, containing relatively few nonzero Fourier coefficients, and this feature is what enables the application of compressed sensing to image processing. In practice, other transformations of an image, such as wavelet transformations, can produce even greater sparsity than the two-dimensional discrete cosine transform, but we will use discrete cosine transforms in this illustrative example both for their simplicity and because the subsequent chapters of this thesis, which focus on chemical applications of compressed sensing, use the discrete cosine transform.

Because the two-dimensional discrete cosine transform of most images is sparse, it is clear that most images contain fewer independent pieces of "information" than their number of pixels. Nevertheless, the dominant paradigm in image acquisition has been "collect, then compress"—in other words, a large number of pixels is acquired, and those pixels are subsequently compressed to be stored. Compressed sensing represents a dramatic change in this paradigm: only a much smaller number of pixels is collected, and the full image is recovered by solving the ℓ_1 optimization problem in equation (1.7). The number of pixels which must be acquired to fully-resolve the image is proportional to that image's information content in the sparse representation, that is, to the number of non-zero coefficients in the two-dimensional discrete cosine transform of the image, as per equation (1.11).

We can represent an image as a vector $\mathbf{b}$ by stacking all of its pixels into a single column. Similarly, we can represent the two-dimensional discrete cosine transform of the

image as a vector **x** by stacking all of the coefficients into a single column. Finally, the matrix A is taken to represent the two-dimensional (inverse) discrete cosine transform which converts the vector **x** into the vector **b**. Since the vector **x** is sparse, we can recover it by sampling only *some* of the pixels of the image to form the vector **b** and then solving the compressed sensing optimization problem in equation (1.7). After the sparse vector **x** has been recovered, it is straightforward to apply the two-dimensional discrete cosine transform to visualize the recovered image.

This procedure is illustrated in Figure 1.5. The bottom-right panel depicts the true exact image used for the example and the top-left panel illustrates the pixels from this image which were sampled to form the measurement vector **b**. As described in the figure, only one-quarter of the pixels were sampled. The top-right panel illustrates the recovered discrete cosine transform from the vector **x** obtained by solving optimization problem (1.7); as expected, this quantity is rather sparse. The bottom-left panel illustrates the recovered image after the (inverse) discrete cosine transform has been applied to the top-right panel. Comparing the reconstructed image to the exact image reveals that the key features of the image are recovered accurately through the compressed sensing process, but that only one-quarter of the pixels were actually needed to recover these key features.

Figure 1.6 illustrates the same process, but with sampling only one-tenth of the pixels in the true image. Again, the key features of the image are recovered but this time the details are not as clear. This situation corresponds to the recovery of a discrete cosine transform that is *too sparse* compared to the true discrete cosine transform of the original image. Even so, the recovered image is a reasonable approximation to the true image. More accurate recovery with fewer pixels may be obtained by choosing a more effective transformation than the discrete cosine transform, one which yields a quantity that is even
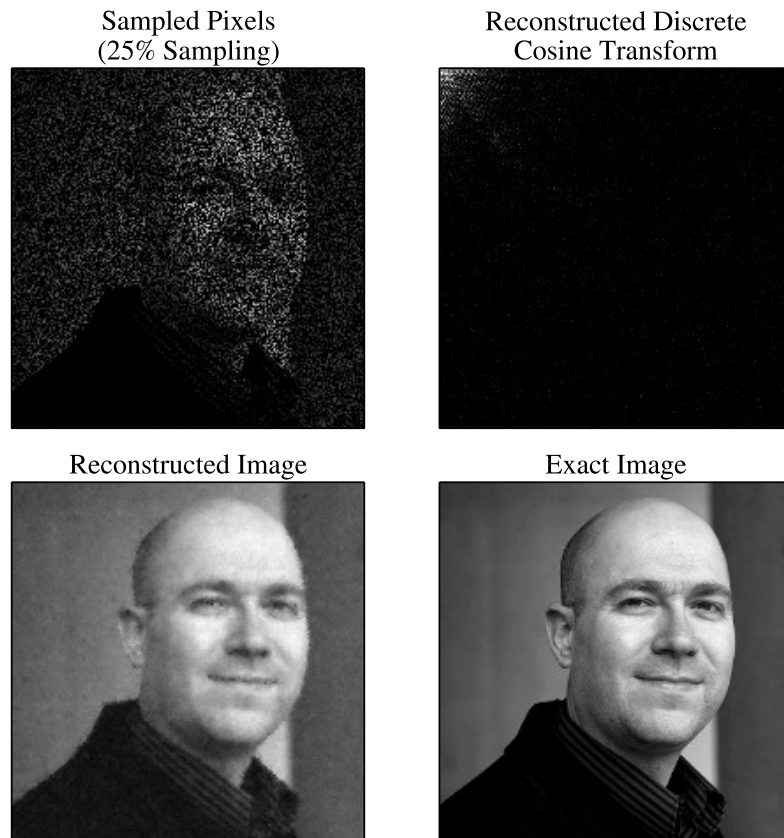
Figure 1.5: This figure illustrates the application of compressed sensing to reconstruct an image by sampling only 25% of its pixels. The sparse discrete cosine transform is recovered from the sampled pixels by solving optimization problem (1.7). The key features of the image are recovered with reasonably high accuracy.
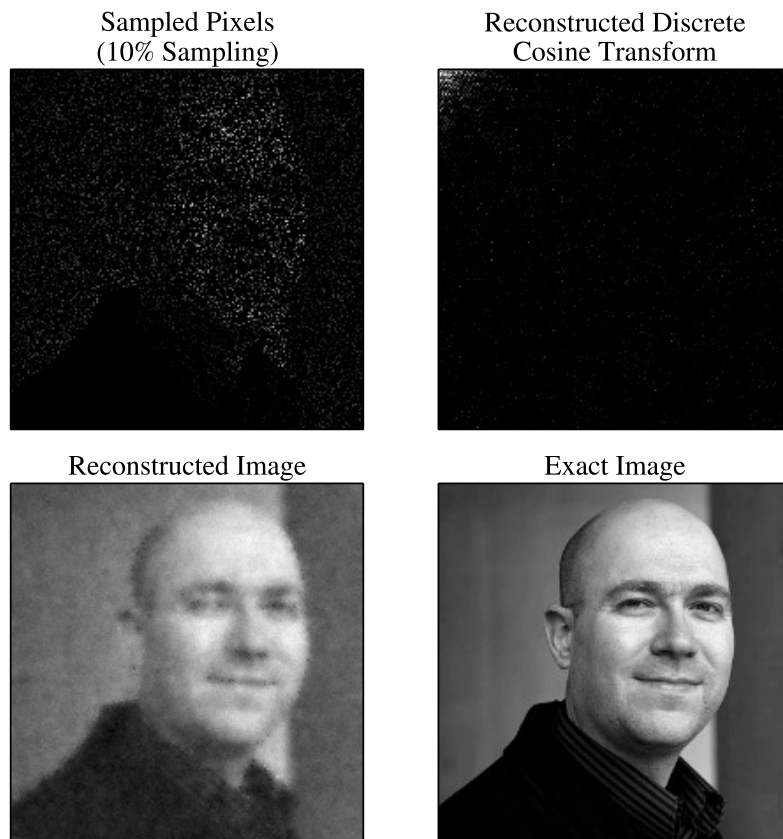
Figure 1.6: This figure illustrates the application of compressed sensing to reconstruct an image by sampling only 10% of its pixels. This time, the accuracy is not as satisfactory.

more sparse, such as a wavelet transformation.

1.3 Applications to Chemistry

The remainder of this thesis contains original applications of compressed sensing to problems in chemistry. In all of these applications, compressed sensing reduces the number of measurements or computations required to obtain a quantity of chemical interest.

Chapter 2 focuses on the application of compressed sensing to accelerate the computation of vibrational and electronic spectra from real-time time-dependent density functional

theory (TDDFT) simulations [13]. Because the molecular systems being simulated have discrete energy levels, the spectra are naturally sparse in the frequency domain. While the typical approach involves real-time propagation in the time domain, followed by application of a discrete Fourier transform to obtain the spectrum, compressed sensing may be used as a drop-in replacement for the discrete Fourier transform to reduce the total propagation time, and hence the computational cost, by a factor of five. Examples include the vibrational spectrum of benzene, the electronic absorption spectrum of benzene, and the circular dichroism spectrum of (R)-methyloxirane, all of which are substantially accelerated by the use of compressed sensing.

Chapter 3 extends the compressed sensing approach to multidimensional spectroscopy and demonstrates the feasibility of its application to experimental data [155]. For the model experimental system of atomic rubidium vapor, which has a sparse two-dimensional electronic spectrum, compressed sensing allows for random undersampling of the time-domain experimental data, down to less than 5% of the experimental dataset, with essentially no loss in spectral resolution.

Chapter 4 applies super-resolution, a technique closely related to compressed sensing, to the simulation of realistic environments for open quantum systems, focusing on the exciton dynamics of the Fenna-Matthews-Olsen (FMO) light-harvesting complex [115]. Realistic models of the environment are challenging to construct because they incorporate many degrees of freedom, but increasing computational resources have enabled their construction from molecular dynamics (MD) simulations combined with time-dependent density functional theory (TDDFT). In this chapter, super-resolution is employed to derive more accurate spectral densities for the FMO complex while reducing the required simulation time by a factor of five. Another attractive feature is that super-resolution directly yields

the parameters of the spectral densities needed for the subsequent simulation of the FMO complex via quantum master equation approaches, without the need for a separate step to extract the parameters from the spectral densities.

Chapter 5 moves away from spectroscopy and presents a new application of compressed sensing to compute matrices whenever the determination of the entries of a matrix constitutes the computational bottleneck [154]. The method is applied to second-derivative Hessian matrices frequently computed in electronic structure calculations to obtain the vibrations of a molecule. Moreover, the method provides a general framework for bootstrapping cheap low-accuracy calculations in order to reduce the required number of expensive high-accuracy calculations; the number of required high-accuracy calculations is related to how much new information they provide over the low-accuracy calculations. When applied to the Hessian matrix of anthracene, this technique results in a threefold speed-up in the calculation, with greater speed-ups possible for larger molecules.

Chapters 6 and 7 describe the incorporation of this compressed sensing technique into the commercial electronic-structure package known as Q-Chem. Chapter 6 describes the implementation for so-called mixed derivatives, where gradients are computed analytically and second derivatives are computed via finite differences. In this case, the incorporation of compressed sensing reduces by roughly half the number of gradients that must be calculated, which translates directly into a reduction in the number of CPU hours required to compute a Hessian matrix. Chapter 7 describes the implementation for fully analytic derivatives, where a savings is possible in the number of first derivatives of the molecular orbital coefficients that must be computed, but where further work remains to translate this savings into a reduction of CPU hours.

Part I

Compressed Sensing for

Spectroscopy

Chapter 2

Application of Compressed Sensing to the

Simulation of Atomic Systems

Apart from minor modifications, this chapter originally appeared as [13]:

"Application of Compressed Sensing to the Simulation of Atomic Systems." Xavier Andrade, Jacob N. Sanders, and Alán Aspuru-Guzik. *Proceedings of the National Academy of Sciences USA*. Volume 109, Number 35: 13928–13933. 2012.

Abstract

Compressed sensing is a method that allows a significant reduction in the number of samples required for accurate measurements in many applications in experimental sciences and engineering. In this work, we show that compressed sensing can also be used to speed up numerical simulations. We apply compressed sensing to extract information from the real-time simulation of atomic and molecular systems, including electronic and nuclear dynamics. We find that, compared to the standard discrete Fourier transform approach, for the calculation of vibrational and optical spectra the total propagation time, and hence the computational cost, can be reduced by approximately a factor of five.

2.1 Introduction

A recent development in the field of data analysis is the compressed sensing (CS) (or compressive sampling) method [34, 57]. The foundation of the method is the concept of *sparsity*: a signal expanded in a certain basis is said to be sparse when most of the expansion coefficients are zero. This extra information can be used by the CS method to significantly reduce the number of measurements needed to reconstruct a signal. CS has been successfully applied to data acquisition in many different areas [17], including the improvement of the resolution of medical magnetic-resonance imaging [111] and the experimental study of atomic and quantum systems [159, 9, 91].

In this article we show that CS can also be an invaluable tool for some numerical simulations with a considerable reduction of the computational cost. We focus on atomistic simulations of nanoscopic systems by using CS to extract frequency-resolved information from real-time methods such as molecular dynamics (MD) and real-time electron dynamics.

MD [146, 5] is one of the most widely used methods to study atomistic systems computationally as it can be used to compute many static and dynamical properties. In MD the trajectory of the atomic nuclei is obtained by integrating their equations of motion either with parametrized force-fields or else by explicitly modeling the electrons [118]. Given its importance, developing methods that can improve the precision and reduce the computational cost of this method, especially for ab initio MD, can have a large impact in the field of atomistic simulation.

Real-time electron dynamics, in particular real-time time-dependent density functional theory (TDDFT) [153], plays a similarly important role in the study of linear and non-linear electronic properties [194, 42, 41, 175]. Due to its scalability and parallelizability, real-time TDDFT is particularly efficient for large electronic systems [10], so an additional reduction

in the computational cost can extend the boundaries of the system-sizes that can be studied.

Many physical properties are represented by frequency-dependent quantities. To obtain these from real-time information, usually a discrete Fourier transform (FT) is used. Our approach is to replace this FT by a calculation of the Fourier coefficients based on the CS method. To obtain a given frequency resolution, the CS method requires a total propagation time that is several times smaller than that required by a FT.

CS has the potential to provide *across-the-board* speedup for many applications involving the computation of sparse spectra. Moreover, this speedup may be obtained without making *any* changes to the underlying propagation code used in different types of electronic and nuclear calculations; one simply replaces the FT algorithm with the CS method, making the approach quite straightforward to implement. This paper introduces CS and demonstrates its broad utility in computational chemistry and physics by applying it to the calculation of various nuclear and electronic spectra of small molecules. The resulting computer code is available as open-source software.

The article is structured as follows. We first introduce the CS method and show how it may be applied to the determination of Fourier coefficients. Next, we apply CS to the calculation of vibrational, optical absorption, and circular dichroism spectra. We then proceed to a discussion of the numerical methods used in our CS implementation. Finally, we offer conclusions and an outlook.

2.2   Compressed sensing

In this section, we briefly introduce the application of the CS method to the calculation of Fourier coefficients. More details about the method and its origins may be found in Refs. [16, 35, 45].

For simplicity, we assume that we want to calculate a certain frequency-resolved quantity $g(\omega)$ that is given by the *sine* transform of a certain time-resolved function $h(t)$

$$g(\omega) = \int_{-\infty}^{\infty} dt \, \sin(\omega t) \, h(t) \tag{2.1}$$

(the analysis is equally valid for the cosine transform). Since we are interested in numerical solutions, we need to think in terms of discrete quantities. For the numerical problem, we consider that $h(t)$ has been sampled at N_t equidistant times $t_j = \Delta t j$. We want to obtain $\{g_1, g_2, \ldots, g_{N_\omega}\}$ at N_ω equally-spaced frequencies $\omega_j = \Delta \omega \, j$.

In principle, the g_k can be directly obtained using the discrete FT,

$$g_k = \sum_{j=1}^{N_t} \Delta t \sin(\omega_k t_j) \, h_j \, . \tag{2.2}$$

However, if we expect that many of the Fourier coefficients are zero, a property known as *sparsity*, we can use this additional information to obtain more precise results. This is the basis for the CS scheme.

We start by reformulating the discrete Fourier transform in equation (3.2) as a matrix inversion problem. From this perspective, we are trying to solve the linear equation for $\boldsymbol{g}$,

$$F\boldsymbol{g} = \boldsymbol{h} \, , \tag{2.3}$$

where F is the $N_\omega \times N_t$ Fourier matrix with entries

$$F_{jk} = \frac{2}{\pi} \Delta \omega \sin(\omega_j t_k) \, . \tag{2.4}$$

Our objective is to obtain sensible results with N_t as small as possible. Thus, we are interested in the case $N_\omega > N_t$, where the linear system is under-determined, and there are many solutions for $\boldsymbol{g}$ (in fact, one of them will be given by equation (3.2)). From all the solutions of equation (3.3), we select the one that has the largest number of zero coefficients:

the sparsest solution. This turns out to be equivalent to the so-called basis-pursuit (BP) optimization problem [35]

$$\min_{\boldsymbol{g}} |\boldsymbol{g}|_1 \quad \text{subject to} \quad F\boldsymbol{g} = \boldsymbol{h} , \tag{2.5}$$

which is what one solves in practice (where $|g|_1 = \sum_k |g_k|$ is the standard 1-norm).

The CS scheme can be generalized to allow for a certain amount of noise in the time-resolved signal. In this case the problem to be solved is known as basis-pursuit de-noising (BPDN)

$$\min_{\boldsymbol{g}} |\boldsymbol{g}|_1 \quad \text{subject to} \quad |F\boldsymbol{g} - \boldsymbol{h}| < \eta , \tag{2.6}$$

where η represents the level of noise in the signal. This is the formulation we use in our case, since we expect a certain amount of noise coming from the finite-precision numerical calculations (and possibly other sources).

2.3 Vibrational spectra

MD can be used to obtain information about the vibrational modes of atomic systems. Experimentally, the quantities that usually give access to the vibrational modes are the infrared and Raman spectra that can be obtained from MD as the Fourier components of the electronic polarization and polarizability, respectively. If we are only interested in the vibrational frequencies, from the nuclear velocities $\{\boldsymbol{v}_i\}$ we can calculate the velocity autocorrelation function

$$\gamma(t) = \frac{\langle \sum_i \boldsymbol{v}_i(t) \cdot \boldsymbol{v}_i(0) \rangle}{\langle \sum_i \boldsymbol{v}_i(0) \cdot \boldsymbol{v}_i(0) \rangle} , \tag{2.7}$$

where the brackets indicate the average over the ensamble. The cosine transform of γ is the vibrational frequency distribution [55]

$$f(\omega) = \int \mathrm{d}t \, \gamma(t) \, cos(\omega t) . \tag{2.8}$$

20

Since this spectrum is composed of a finite number of frequencies (less than three times the number of atoms in the system), the calculation is ideal for the CS method.

To illustrate the properties of the CS method we start with a simple case, the single vibrational frequency of a diatomic molecule, Na_2, which we simulate using ab-initio molecular dynamics. In Figure 2.1, we show how the vibrational spectrum depends on the amount of time for which the velocity autocorrelation is calculated, both for CS and for the discrete FT with a standard polynomial damping function. While the FT requires long propagations to resolve the vibrational frequency, the CS method gives a well-defined peak even with less than one oscillation of the molecular vibrational mode. That the peak is well defined, however, does not imply that the peak position is converged. As seen in Figure 2.2, the peak position oscillates with the total sampling time until it converges to the true value after a few periods are sampled. For the FT, on the other hand, the peak position, obtained by fitting the spectrum to a Gaussian function, is well defined when more than one period is sampled, but the width of the peak converges very slowly with the sampling time. This difference in behavior of CS and FT has been reported previously, for example, in Ref. [205]. We remark that the CS process does not use any additional information about the signal beyond assuming it is sparse.

The advantage of the CS approach is further illustrated when several closely-spaced frequencies must be resolved. To demonstrate this, we calculate the vibrational spectrum for a benzene molecule from an ab initio MD simulation in Figure 2.3. We can see that the CS approach with 1000 fs gives a spectrum that is better resolved than the FT results for 5000 fs. This is directly translated into a reduction of the computational time by five times or more. It is reasonable to expect that equivalent gains can be obtained for the computer simulation of other vibrational spectroscopies such as infrared and Raman.
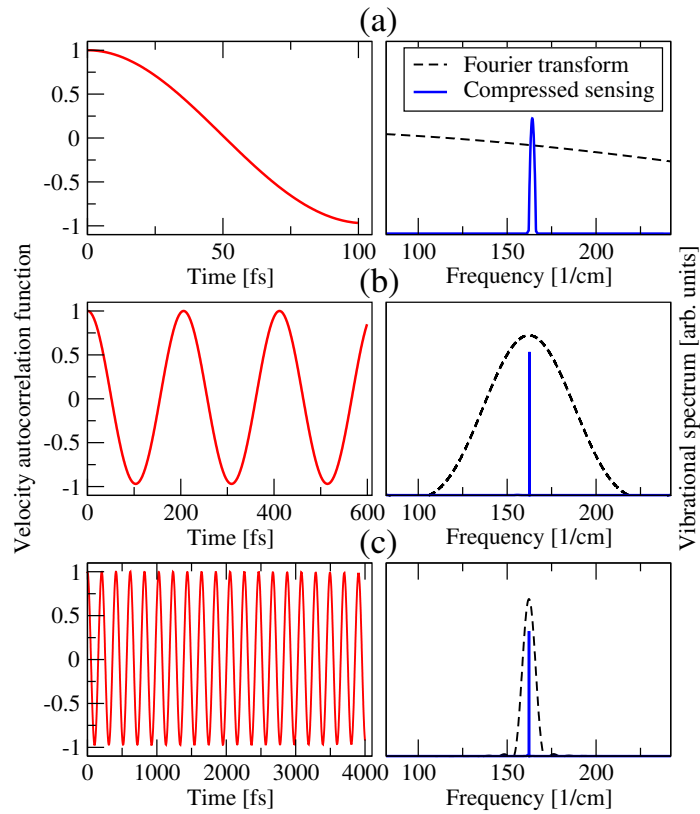
Figure 2.1: Frequency distribution spectrum of Na_2 calculated using a Fourier transform and compressed sensing for different total propagation times: (a) 100 fs, (b) 600 fs, and (c) 4000 fs. The left plots show the velocity autocorrelation function and the right plots show the frequency spectrum.
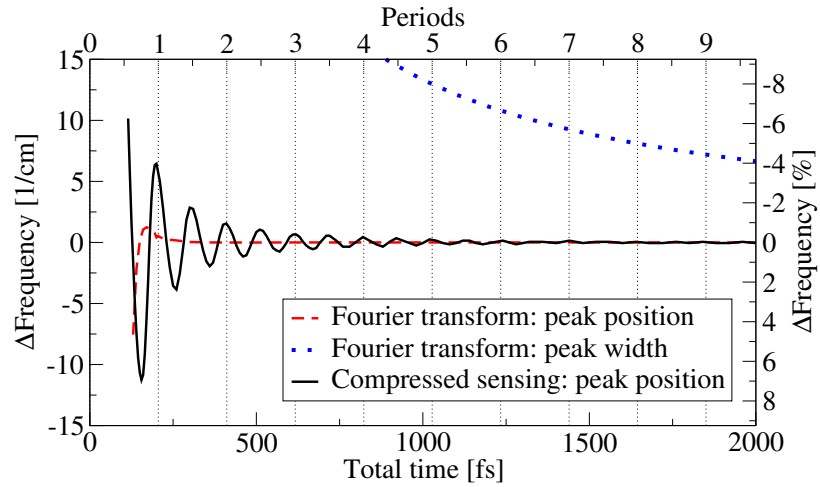
Figure 2.2: Error in the vibrational frequency of Na_2 computed by compressed sensing with respect to total time. For comparison we plot the positions and width of the peak obtained by using a discrete Fourier transform. For the Fourier transform, the position, ω_0, and the width, σ, are fit by assuming the peak has a Gaussian form $A \exp[-(\omega - \omega_0)^2/(2\sigma^2)]$. For compressed sensing, the peak width corresponds to the energy spacing, 0.01 1/cm.
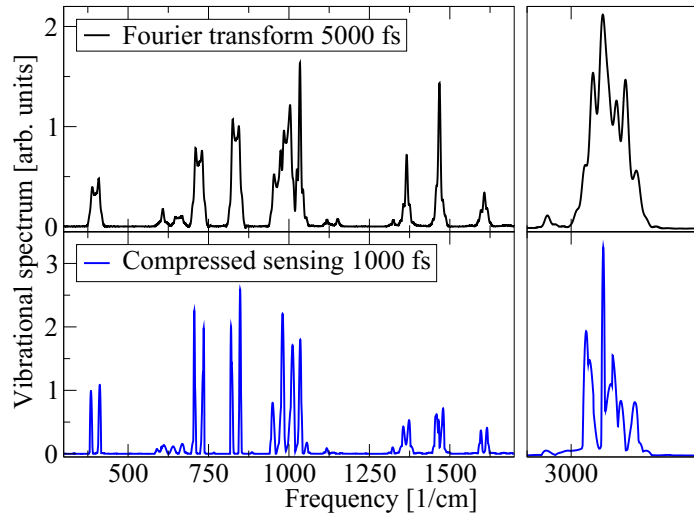
Figure 2.3: Vibrational frequency distribution spectrum of benzene. Comparison of a compressed sensing calculation with 1000 fs and a Fourier transform with 5000 fs.

2.4 Optical absorption spectra

Optical absorption is an electronic process. While it can be calculated from a linear response framework [39, 11], it can also be obtained from real-time electron dynamics [194]. To obtain the spectrum from real-time dynamics the electronic system is propagated under the effect on an electric field of the form $\boldsymbol{E}(\boldsymbol{r}, t) = \boldsymbol{\kappa}\delta(t)$. From the propagation the time-dependent dipole moment $\boldsymbol{p}(t)$ is obtained, and from the dipole the frequency-dependent polarizability can be obtained as (atomic units are used in the next two sections)

$$\alpha_{ij}(\omega) = \frac{1}{\kappa_i} \int_0^\infty dt\, e^{-i\omega t}\, [p_j(t) - p_j(0)] \ . \tag{2.9}$$

In order to obtain the full $\boldsymbol{\alpha}$ tensor three propagations are required (with $\boldsymbol{\kappa}$ in different directions).

The absorption cross-section is related to the trace of the imaginary part of the polarizability tensor

$$\sigma(\omega) = \frac{4\pi\omega}{3c}\mathfrak{Im} \sum_{i=1}^{3} \alpha_{ii}(\omega) \ . \tag{2.10}$$

The optical absorption spectrum is an ideal candidate for the application of CS. For a molecule, the electronic transitions between bound states produce a discrete spectrum in the low energy region. At higher energies, the transitions to unbound states produce a continuous spectrum. In general, electronic structure calculations cannot capture this continuous spectrum and approximate it as a sequence of discrete excitations, so in practice the whole spectrum is sparse.

In Figure 2.4, we show the optical absorption spectrum for benzene calculated via real-time TDDFT. There we illustrate the effect of the propagation time on the spectrum for CS and FT. From the figure, it is clear that the CS method is capable of resolving the spectrum much better and with a shorter propagation time than a discrete FT. Since the
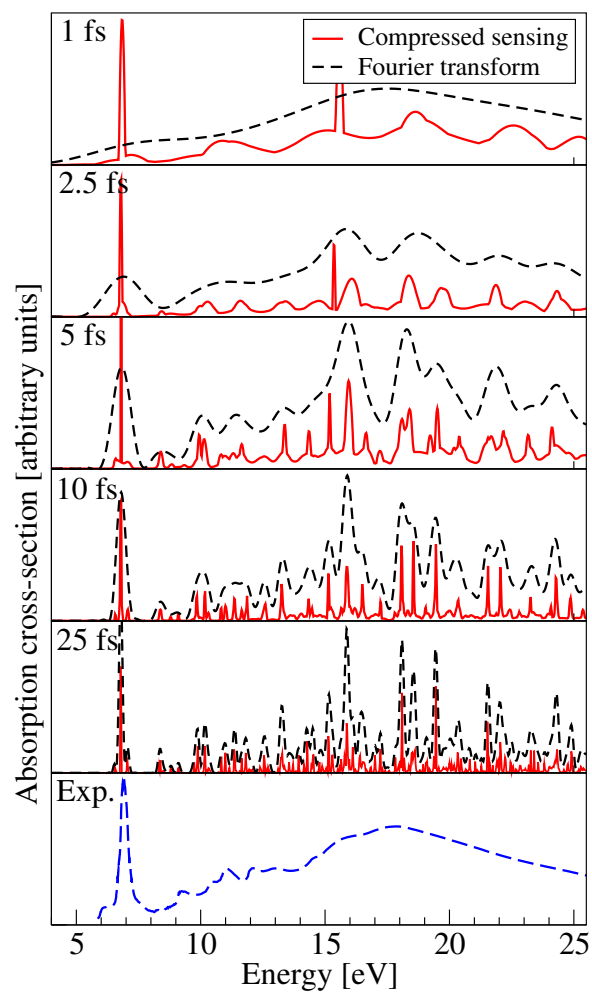
Figure 2.4: Optical absorption spectra for benzene computed from real-time TDDFT with different propagation times. Comparison between compressed sensing and discrete Fourier transform. Experimental results from Ref. [95].

two methods have different convergence properties, as discussed in the previous section, it is difficult to rigorously quantify the speed-up afforded by CS. Nevertheless, by comparing the FT at 25 fs with CS at 5 fs, we can see that the FT requires approximately five times the propagation time as CS to achieve a similar resolution. Note that for a sampling time of 1 fs, the CS spectrum is not particularly sparse; we attribute this to the fact that perhaps the 1-norm minimization is not a good measure of sparsity in the very short-time regime [34].

2.5 Circular dichroism spectra

Another property that can be calculated from real-time electron dynamics is circular dichroism (CD) spectra [195, 185]. A CD spectrum measures the difference in a chiral molecule's response to left and right circularly-polarized light. The real-time calculation is performed in the same manner as the optical absorption case, but the key quantity to be calculated is now the time-dependent orbital magnetization $\boldsymbol{m}(t)$. The frequency-dependent electric-magnetic cross-response tensor may be obtained as

$$\beta_{ij}(\omega) = \frac{ic}{\omega \kappa_i} \int_0^\infty dt \, e^{-i\omega t} \left[m_j(t) - m_j(0) \right] \ . \tag{2.11}$$

The rotatory strength, which is the quantity typically plotted in CD spectra, is related to the trace of the imaginary part of $\boldsymbol{\beta}(\omega)$ according to

$$R(\omega) = \frac{\omega}{\pi c} \mathfrak{Im} \sum_{i=1}^3 \beta_{ii}(\omega) \ . \tag{2.12}$$

The rotatory strength $R(\omega)$ is suitable to the CS scheme because it is sparse in frequency space. In fact, the peaks in a CD spectrum are located at the same positions as in an absorption spectrum. However, the CD spectrum contains both positive and negative peaks.

Figure 2.5 compares the CD spectrum for (R)-methyloxirane as computed by FT and CS for two different propagation times (10 fs and 50 fs). As can be seen from the figure, for a

given propagation time, the CS method provides better spectral resolution than the discrete FT. In fact, just as with linear absorption, FT requires a propagation time approximately 5 times as long as CS to obtain a comparable spectral resolution (as can be seen by comparing the 50 fs FT with the 10 fs CS).

Figure 2.5 also illustrates another feature of CS: unlike a direct FT, the CS method is non-linear. Adding together time-resolved signals, then applying CS, generally gives different results from applying CS first and then adding together the results in the frequency domain; this is particularly the case if not all the peaks are well-resolved. In other words, the use of CS to convert time-resolved data into the frequency-domain, as in equation (2.11), and the calculation of the trace, as in equation (2.12), do not commute. Hence, there are two approaches to obtain the CD spectrum: we can perform CS for each propagation direction and then compute the trace, or we can compute the trace in the time-domain and then perform CS. Both approaches are shown in Figure 2.5; at 10 fs, they give similar but not identical results. Performing the CS after the trace appears to give a cleaner spectrum with fewer and better-resolved peaks. This is to be expected from the sparsity requirement. For longer propagation times (50 fs), all of the peaks are more fully resolved and the two approaches converge. In any event, both approaches to CS provide much improved resolution over a direct FT for a given propagation time. This effect is also present in the absorption spectra, but to a much lesser degree, probably because the spectra are non-negative.
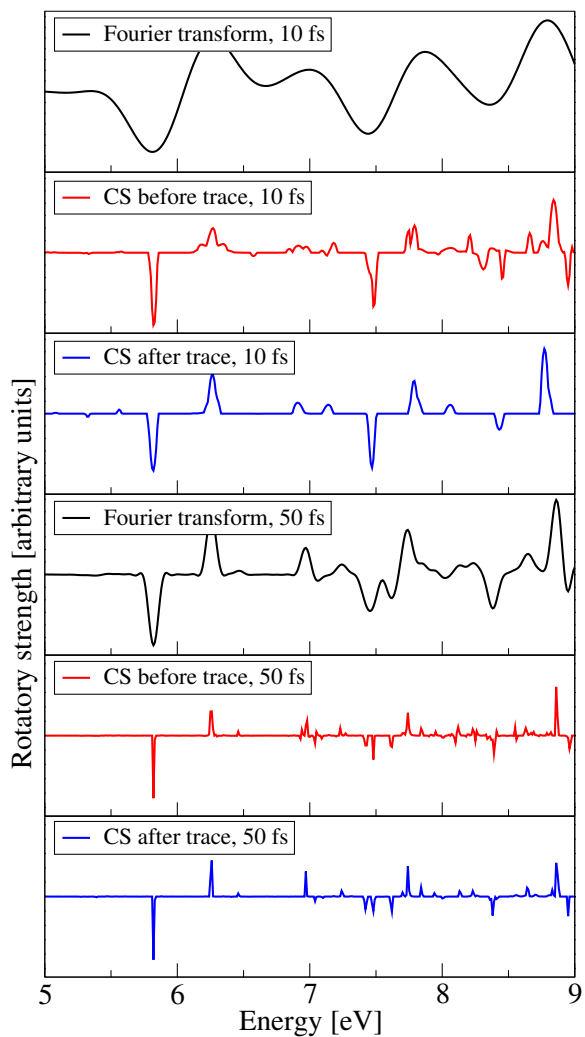
Figure 2.5: Circular dichroism spectrum computed for (R)-methyloxirane from real-time TDDFT with different propagation times. Comparison between discrete Fourier transform and compressed sensing (CS). Since the CS process is non-linear we compute the spectra in two ways. CS before trace: the spectra is calculated for each direction using CS and then the trace, equation (2.12), is computed. CS after trace: the trace is calculated in the time domain and then the CS process is used over this averaged signal.

2.6   Numerical methods

Numerically, to find a spectrum using the CS method we need to solve equation (5.2). This is not a trivial problem, so we rely on the SPGL1 algorithm developed by van den Berg and Friedlander [184]. To avoid numerical stability issues we work with a normalized BPDN problem, where the prefactor $2\Delta\omega/\pi$ of the F matrix, equation (3.4), is left out and $\boldsymbol{h}$ is normalized. These prefactors are then included in $\boldsymbol{g}$ after the solution is found. This has the additional advantage of making the noise parameter η of equation (5.2) dimensionless.

Since we do not a have an *a priori* estimate for η, we do not set it directly. As the SPGL1 algorithm finds a sequence of approximate solutions with decreasing values of η, we set the target value to zero. We assume that the calculation is converged when the value of η falls below a certain threshold (10^{-7}) or the active space of the system, the set of non-zero coefficients, has not changed for a certain number of iterations (50). In the former case we consider that a solution of the BP problem, equation (2.5), has been found. For all the calculations presented here $\eta < 10^{-3}$.

CS is much more costly numerically than the discrete FT approach, as it usually involves several hundreds of matrix multiplications. However, this is not a problem for the applications we are proposing since this process normally only takes a few minutes, much less than the computation time required to simulate the real-time dynamics of large atomic systems.

All the calculations presented in this article were performed using the OCTOPUS code [40, 10] at the (time-dependent) density functional theory level with the PBE exchange correlation functional [137]. The adiabatic molecular dynamics calculations were performed from first principles using the modified Ehrenfest method [8, 12] with an ionic time-scale factor (μ) of 30 for Na_2 and 5 for benzene. The systems were given initial velocities equivalent to

300 K and the MD is performed at constant energy.

All calculations used norm-conserving pseudo-potentials with a real-space grid discretization. The shape of the grid is a union of boxes around each atom. We chose the grid parameters to ensure proper convergence of the results to an accuracy of 1%. For Na_2 we use a spacing of 0.375 a.u. with a sphere radius of 12 a.u., and the MD time-step is 0.057 fs. For benzene, the grid-spacing is 0.35 a.u., the radius is 14 a.u., and the time-step is 0.0085 fs for MD and 0.0017 fs for real-time TDDFT. For (R)-methyloxirane, the spacing is 0.378 a.u., the sphere radius is 15.1 a.u., and the time-step is 0.0008 fs for real-time TDDFT. For the vibrational spectrum calculation we use a time-step 10 times the one of the MD, the energy step is 0.01 1/cm, and the maximum spectrum energy is 5000 1/cm. For the benzene optical absorption spectra, we use a time-step of 0.0017 fs, the energy step is 0.027 eV, and the maximum spectrum energy is 820 eV. For the (R)-methyloxirane circular dichroism spectra, the time-step is 0.0008 fs, the energy step is 0.01 eV, and the maximum spectrum energy is 330 eV. The structure of benzene was taken from Ref. [53] and the structure of (R)-methyloxirane was taken from Ref. [37].

All discrete FTs were performed using third-order polynomial damping: each signal at time t was multiplied by $p(t) = 1 - 3(t/T)^2 + 2(t/T)^3$ prior to Fourier transform, where T is the time-length of the signal [196].

The SPGL1 method used for CS was implemented into OCTOPUS based on a Fortran translation of the original Matlab code of van den Berg and Friedlander [184]. We plan to release this implementation as a standalone tool in the near future (for the moment the code can be obtained from the OCTOPUS repository).

2.7 Conclusions

We have shown that the CS method can be applied to the numerical calculation of different kinds of atomic and electronic spectra. This results in a significant reduction of the computational time required for the numerical simulations. For the systems tested in this work, we find that is possible to obtain an speedup of five for CS as compared to the discrete FT. Since the number of samples required for CS depends on the sparsity of the signal [35], this ratio could change for other systems.

The effect of this reduction is to increase the size of the systems that are currently accessible to numerical simulations, and to make possible simulations with more precise, but more costly, methods. It also means that other types of simulations become more affordable from a real-time perspective, for example the combined dynamics of nuclei and electrons that are constrained to short simulation times by the fast dynamics of the electron.

In this work, we have shown the application of CS to the calculation of a few types of spectra, but the method most likely can be applied to other quantities as well, such as non-linear optical response [41, 175], magnetic circular dichroism [109], semi-classical nuclear dynamics [44], 2D spectroscopy [114, 198], etc. Of course, the method is not limited to atomistic simulations and could be applied to simulations in all scientific fields.

Probably the main advantage of CS is that it is a simple and universal approach that works as a drop-in replacement for the FT for many applications where discrete or sparse spectra need to be calculated. The main limitation of the CS approach is that it may not be as beneficial for quantities that are not sparse.

There are some cases where the sparsity requirement might be circumvented. For example, though the real part of the polarizability tensor is not sparse, it could be computed from the imaginary part by using the Kramers-Kronig relation. Another example is crys-

talline systems [25], where there is a continuum of excitation energies. In this case it might be possible to apply the CS scheme to each k-point separately.

Another important issue is the performance of CS in resolving spectral features such as peak broadening. These features do not appear in the spectra calculated in this article, but might appear when internal structure or environmental effects are considered. This will be the focus of future work. Another interesting continuation of our work would be to compare CS with other approaches for calculating spectral properties such as the maximum entropy method [70]. This method is related to CS but may better reproduce broadened features in spectra [171].

Nevertheless, we want to emphasize that for the sparse spectra obtained in atomistic simulations, compressed sensing is a promising method. We expect that compressed sensing will become widely used in the scientific computing community once its advantageous properties become more widely known. The main difficulty in the adoption of CS is that it is more complex to implement than a discrete FT. This problem can be solved by providing libraries and utilities that can be used by researchers. However, the CS method also has features researchers will need experience to understand. For example, the CS method is non-linear as discussed and the peak width is not always related to the convergence of the spectrum.

We believe that the direct application of the compressed sensing methodology to numerical simulation opens the path for more challenging applications. The principles of sparsity could be used to design algorithms for numerical simulations that have a reduced computational cost not only in the number of operations, but also in memory and data transfer bandwidth requirements; we could call this "compressed computing."

Acknowledgments

Chapter 3

Compressed Sensing for Multidimensional

Spectroscopy Experiments

Apart from minor modifications, this chapter originally appeared as [155]:

Abstract

Compressed sensing is a processing method that significantly reduces the number of measurements needed to accurately resolve signals in many fields of science and engineering. We develop a two-dimensional variant of compressed sensing for multidimensional spectroscopy and apply it to experimental data. For the model system of atomic rubidium vapor, we find that compressed sensing provides an order-of-magnitude (about tenfold) improvement in spectral resolution along each dimension, as compared to a conventional discrete Fourier transform, using the same data set. More attractive is that compressed sensing allows for random undersampling of the experimental data, down to less than 5% of the experimental dataset, with essentially no loss in spectral resolution. We believe that by combining powerful resolution with ease of use, compressed sensing can be a powerful

tool for the analysis and interpretation of ultrafast spectroscopy data.

3.1 Introduction

Compressed sensing (CS) is a state-of-the-art signal processing method which has recently become popular throughout the physical and biological sciences. The method is founded on the concept of *sparsity*. When a signal is known to be sparse in a certain basis (i.e. most of the coefficients are negligibly small), this knowledge can be used to dramatically reduce the number of measurements required to reconstruct the signal [34, 57]. This method has been applied to many areas of research, ranging from magnetic resonance imaging [111] to superresolved imaging of single molecules [205] and quantum process tomography [159].

In this letter, we present the application of CS to experimental ultrafast two-dimensional (2D) optical spectroscopy. Multidimensional spectroscopy [79, 71, 93, 88] is an important tool for studying ultrafast dynamical processes in complex molecular systems. For instance, it can be used to analyze vibrational energy transfer at liquid/air interfaces on picosecond timescales [201] or exciton dynamics in natural light harvesting systems at hundreds of femtoseconds [30, 198, 200]. It can also be applied to the efficient detection and identification of molecules, which is one of the crucial challenges in chemistry, biology, and medicine with important applications to molecular sensing, chemical separation, and DNA analysis [127, 67]. Frequently, in these nonlinear optical techniques, the data is collected in the time domain, and then Fourier transformed to the frequency domain.

While the Fourier transform (FT) is a standard technique in optical spectroscopy, different data processing methods have been applied previously mainly in nuclear magnetic resonance (NMR) spectral analysis [96, 123]. All these techniques can be classified into two main groups: (a) methods requiring periodic sampling, such as the discrete FT [61, 29]

35

and linear prediction (LP) [107, 96] including its variations [176]; and (b) methods that use non-uniform sampling strategies, for instance maximum entropy methods (see refs. in [123]). Formally, the CS method belongs to the latter group, where either random or periodic sampling is applied with no pre-assumptions about the shape of the signal besides its sparsity. The main drawback of periodic sampling methods is the dual constraint on data collection: long time data collection is required to obtain good spectral resolution, while closely-spaced sample points are needed to avoid aliasing. Non-uniform sampling methods can alleviate these issues, but sometimes introduce sampling artifacts [123].

Earlier, some of us showed that CS can also be used to significantly reduce the computational cost of atomistic simulations [13]. In that work, the application of CS was particularized for molecular dynamics and real-time time-dependent density functional theory simulations for obtaining linear spectra (vibrational, optical absorption, and circular dichroism) [10]. However, CS can be applied to other types of simulations and experimental techniques. Multidimensional non-linear spectroscopy is an interesting and relevant candidate to explore the possibilities of CS. Recently, CS has been pursued for the reconstruction of one of the dimensions in 2D NMR data [165] and for the theoretical simulation of 2D spectra [7].

As an illustration we apply the method to an experimental system, atomic Rubidium vapor, which is frequently used as a test model for multidimensional spectroscopy techniques [179, 186, 178]. Many multidimensional experiments are inherently limited in the amount of time-domain data that may be collected, either due to measurement constraints or to more fundamental limitations such as the timescale of the dynamics one wishes to explore (which may be very short due to decoherence and other processes). We believe CS can extend the range of spectroscopically-observable dynamics from limited time domain
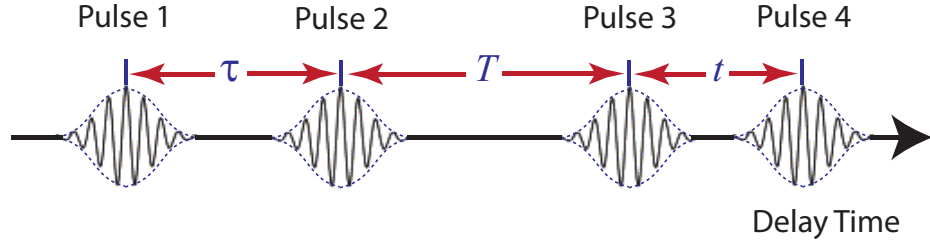
Figure 3.1: Schematic illustration of the sequence of four pulses used in 2D optical spectroscopy.

data. Here, we show that CS resolves spectral lines about an order-of-magnitude better as compared to a discrete FT using the same dataset. While CS can be applied with a uniform sampling strategy, the full potential of the method is realized when a non-uniform random sampling is adopted [57]. In particular, we find that CS with random undersampling down to only 5% of the data yields spectra that are comparable in resolution to those obtained using all of the experimental data.

Another benefit of CS is that it is quite easy for experimentalists to integrate into existing schemes. As discussed below, our basic approach simply replaces the 2D discrete Fourier transform by a new 2D compressed sensing scheme. It is our hope that CS's easy portability will help it become a common method among experimental spectroscopists.

The rest of this letter is structured as follows. We first present the compressed sensing method and outline its application to the resolution of 2D ultrafast optical spectra. Next, we apply the method to a model experimental system, namely gas-phase Rubidium atoms, and show how CS can be used to obtain a tenfold improvement in spectral resolution as compared to FT. We then turn to the effects of different sampling patterns on signal reconstruction, focusing on how random undersampling can be used without sacrificing spectral resolution. Finally, we offer conclusions and a future outlook.

We begin by describing how the CS method can be applied to 2D ultrafast optical spectroscopy; see refs. [16, 35, 45] for more detailed information about CS. As an illustration, we focus on four-wave mixing experiments, similar to those used to study coherent energy transfer in light-harvesting complexes, quantum dots, and other systems of physical and biological interest. Typically, these experiments involve irradiating a sample with four optical pulses and varying the time gaps between the pulses (Figure 3.1). Defining the time gaps as τ (the coherence time between pulses 1 and 2), T (the population time between pulses 2 and 3), and t (the detection time between pulses 3 and 4), the signal measured in the time domain is a function $h(\tau, T, t)$. The standard approach is to perform a 2D discrete Fourier transform in τ and t to obtain the frequency spectrum $g(\omega, T, w)$

$$g(\omega, T, w) = \int d\tau \, dt \, e^{i\omega\tau} \, e^{iwt} \, h(\tau, T, t) \ . \tag{3.1}$$

This frequency spectrum is typically plotted in ω–w space for different values of T, and the dynamics of the peaks then give information about the dynamics of the underlying system.

It has been shown previously how the CS method can be used to replace a *one-dimensional* discrete FT in converting time-resolved data into the frequency domain [34, 13], obtaining a significantly better frequency resolution. Our basic approach here is to generalize the method to the 2D case. A higher multidimensional CS scheme can also be obtained as a straightforward generalization.

For 2D spectroscopy the measurement is usually done for a finite set of points $\{(\tau_j, t_k)\}$ arranged on a $N_\tau \times N_t$ grid. We call the set of measured time-resolved values $\boldsymbol{h} = \{h_{[jk]}\}$. To obtain the 2D spectra we need to calculate the associated set of Fourier coefficients $\boldsymbol{g} = \{g_{[lm]}\}$ defined on a grid of $N_\omega \times N_w$ points $\{(\omega_l, w_m)\}$ in frequency space. This can

be done by using a discrete FT

$$g_{[lm]} = \sum_{j=1}^{N_\tau} \sum_{k=1}^{N_t} \Delta\tau \Delta t \, e^{i\omega_l \tau_j} \, e^{iw_m t_k} \, h_{[jk]} \,, \qquad (3.2)$$

where $\Delta\tau$ and Δt are the values of the spacing of the 2D time grid.

However, if many of the frequency components are zero, then we only need a number of samples proportional to the number of non-zero coefficients in the spectrum [57]; this is the fundamental principle behind CS. The CS approach imposes sparsity as additional requirement to the determination of the Fourier coefficients, thereby significantly reducing the number of samples required for a well-resolved spectrum. We start by reformulating the 2D discrete Fourier transform in equation (3.2) as a linear problem. From this perspective, we are solving the linear equation for $\boldsymbol{g}$,

$$F\boldsymbol{g} = \boldsymbol{h} \,, \qquad (3.3)$$

where F is the $(N_\omega N_w) \times (N_\tau N_t)$ inverse Fourier matrix with entries

$$F_{[jk][lm]} = \frac{4}{\pi^2} \Delta\omega \Delta w \, e^{-i\omega_l \tau_j} \, e^{-iw_m t_k} \,. \qquad (3.4)$$

Our objective is to obtain sensible results with N_τ and N_t as small as possible, so the linear problem is underdetermined $(N_\omega N_w > N_\tau N_t)$ and as such it has many solutions. From all of them we select the sparsest one: the solution with the largest number of zero coefficients. In practice, this solution can be obtained by solving the basis-pursuit de-noising (BPDN) problem [34]

$$\min_{\boldsymbol{g}} |\boldsymbol{g}|_1 \quad \text{subject to} \quad |F\boldsymbol{g} - \boldsymbol{h}|_2 < \eta \,, \qquad (3.5)$$

where the 1-norm $|\boldsymbol{g}|_1 = \sum_{[jk]} |g_{[jk]}|$ is used. The η value accounts for a certain level of noise that can be present in the experimental data. In all calculations which follow, we set $\eta < 10^{-4}$.
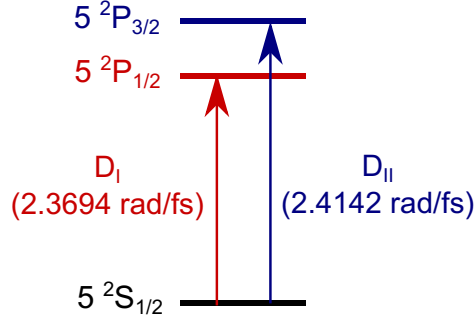
Figure 3.2: Energy level diagram for atomic ^{87}Rb vapor. The transition frequencies are obtained from ref. [170].

As an example, we consider phase-modulation 2D fluorescence spectroscopy (PM-2DFS) data collected from atomic ^{87}Rb vapor [178]. The lowest electronic transitions in ^{87}Rb gas may be considered as a quantum three-level system with ground state $5\ ^2S_{1/2}$, first excited state $5\ ^2P_{1/2}$, and second excited state $5\ ^2P_{3/2}$, as illustrated in Figure ??. The four light pulses are produced by a titanium saphire laser [with full width at half maximum (FWHM) ~ 42 fs] which is resonant with electronic transitions of ^{87}Rb: $5\ ^2S_{1/2} \rightarrow 5\ ^2P_{1/2}$ (with transition frequency 2.3694 rad/fs) and $5\ ^2S_{1/2} \rightarrow 5\ ^2P_{3/2}$ (with transition frequency 2.4142 rad/fs). We considered ^{87}Rb vapor to be an ideal candidate for our 2D CS method as its 2D electronic spectrum is expected to be sparse in frequency space, with diagonal and cross peaks corresponding to the transitions just mentioned. The natural lifetime of electronic excitations in ^{87}Rb is about 25 ns, which corresponds to very narrow natural linewidths (D_I 36.1 and D_{II} 38.1 rad/ms). Full and extensive details of the PM-2DFS experimental method used to collect the data are given in refs. [178, 110, 138]. The experiments yield time-resolved fluorescence-detected "sum" and "difference" signals [$h_{\text{sum}}(\tau, T, t)$ and $h_{\text{diff}}(\tau, T, t)$] which are analogous, respectively, to the nonrephasing and rephasing third-order polarizations collected in more traditional four-wave mixing experiments. In the rest
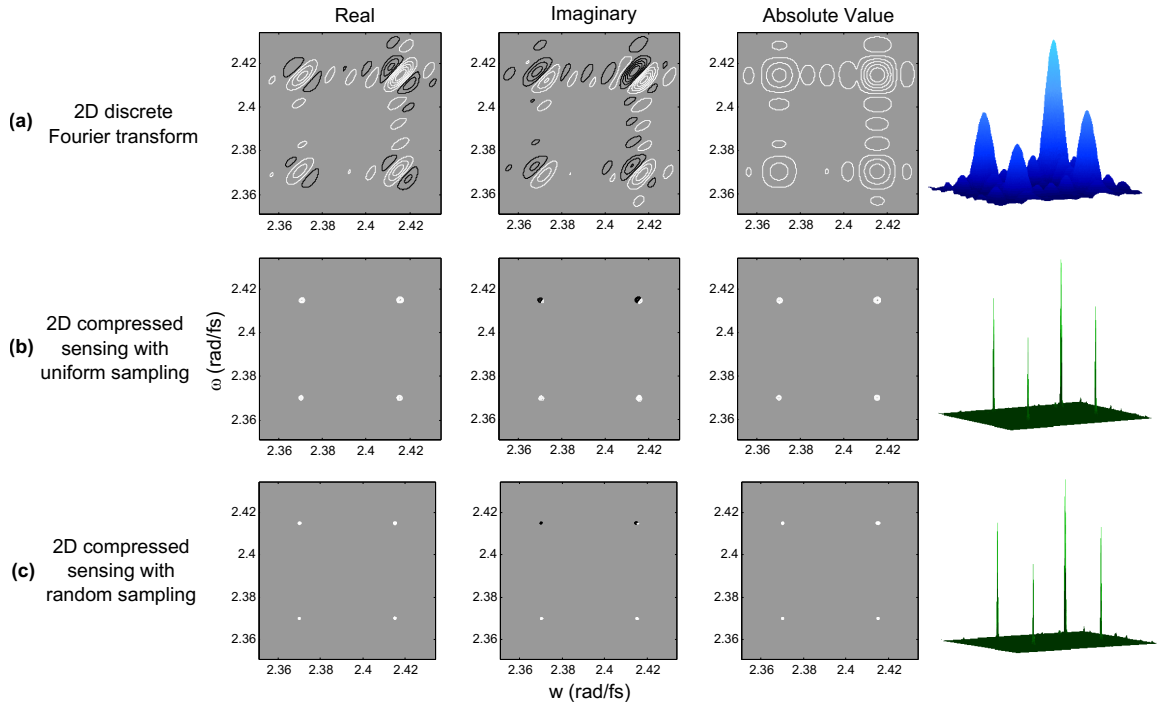
Figure 3.3: Comparison between Fourier transform and compressed sensing, with both uniform and random sampling, in resolving 2D optical spectrum of gas-phase atomic Rb at population time $T = 140$ fs. (a) FT using uniform time grid of 25×25 evenly-spaced points with $\Delta\tau = \Delta t = 26.7$ fs. (b) CS using exactly the same time-domain data as (a). (c) CS using random sampling of 25^2 time points drawn from full grid of 45×45 time points.

of this paper we focus on the "difference" (rephasing) signal obtained for population time $T = 140$ fs. We have observed exactly the same improvements upon applying our 2D CS method to other population times and to "sum" (nonrephasing) signals.

The main results of this letter are summarized in Figures 3.3–3.5, which compare the performance of the 2D discrete FT and our 2D CS method, with both uniform and random sampling, in resolving the spectral peaks corresponding to transitions in the atomic [87]Rb vapor. Several conclusions about 2D CS may be drawn by systematically comparing different parts of the figures.

The first important comparison is between Figure 3.3(a) and Figure 3.3(b). These two figures present the results of applying the 2D FT and 2D CS methods to exactly the same set of time-domain experimental data (a grid of 25 evenly-spaced points in τ and 25 evenly-spaced points in t with $\Delta\tau = \Delta t = 26.7$ fs, for a total sampling time of 667 fs in each time dimension). We see that CS produces peaks that are better resolved in frequency space than those obtained by the discrete FT from exactly the same data. In fact, the CS peaks are better resolved by an order of magnitude (about 10 times) along each dimension, consistent with the results we previously obtained for one-dimensional spectra [13]. In Figure 3.3(c), we compute the 2D spectrum using a randomly selected subset of 25^2 data points from the entire 45×45 grid of available time-domain data. While this is not a full random sampling on the continuous space, it allows us to use the same data to investigate the effects of random sampling in CS.

The advantage of 2D CS over 2D FT is quantified in Figure 3.4. This graph compares the linewidth of the upper-right peak obtained by FT (black) and CS (red) while varying the number of time points sampled in each dimension. Since the density of time points is held constant, this corresponds to varying the total experimental sampling time along the τ and t dimensions. As expected, for longer sampling times, the peaks become narrower and better resolved with both FT and CS methods. However, CS consistently provides an order of magnitude improvement over the discrete FT for a given sampling time. In fact, applying CS to the first 10 time points along each dimension (total sampling time 267 fs) yields higher spectral resolution than applying the discrete FT to 45 time points along each dimension (total sampling time 1201 fs).

In Figure 3.4 we also compare the linewidths generated by the random-sampling CS approach (blue dots). For each point we average over 10 different sets of random samples.

Figure 3.4: Comparison of linewidths obtained using Fourier transform and compressed sensing, with both uniform and random sampling, as the number of time-domain measurements is varied. The discrete FT (black) and the CS with uniform grid sampling (red) are done using the first N^2 evenly-spaced points on a two-dimensional time grid with $\Delta\tau = \Delta t = 26.7$ fs. The CS with random grid sampling (blue) is done using N^2 points drawn randomly from the entire 45×45 time grid. Linewidths are reported as the mean of the widths along the ω and w axes. For each random sampling data point, linewidths are averaged over ten different random samplings; error bars denote the standard deviation.

As shown in the figure, random sampling drastically reduces the number of time-domain measurements needed to obtain well-resolved spectra, even beyond what can be achieved by using CS with uniform sampling. Even randomly sampling only 10^2 points (about 5% of the total number of points) in the time-domain (the far-left blue point), we obtain linewidths that are not much larger than sampling the entire data set of 45^2 points (the far-right red point). In other words, randomly undersampling from the time-domain grid provides nearly the same spectral resolution as if the entire grid was sampled and the spectral resolution does not change much as the extent of random undersampling increases. Note, however, that attempting to sample 5^2 points resulted in failed spectral reconstruction; the extent of undersampling that is permissible is known to scale with the sparsity of the signal to be reconstructed [57]. It is reasonable to expect that a fully random CS approach would provide even better resolution in comparison with selecting random points from a grid as done in these calculations.

Finally, to confirm the accuracy of FT and CS, with both uniform and random sampling, we computed the error in the frequency difference between the two optical transitions (Figure 3.5(a)) and the error in the average frequency of these transitions (Figure 3.5(b)) for all three methods. In all cases, both the average and the difference of the frequencies fall within a few frequency grid points of the true values (the true values are 0.0448 rad/fs for the frequency difference and 2.3918 rad/fs for the frequency average [170]). Moreover, while the uniform CS sampling gives a small improvement over the discrete FT, the random CS sampling provides more accurate frequency differences and average frequencies even using a minimal number of time-domain measurements.

Uniform CS sampling with reduced density of the grid points, for example by sampling every N^{th} time point, results in aliasing once the frequency bandwidth is smaller than the
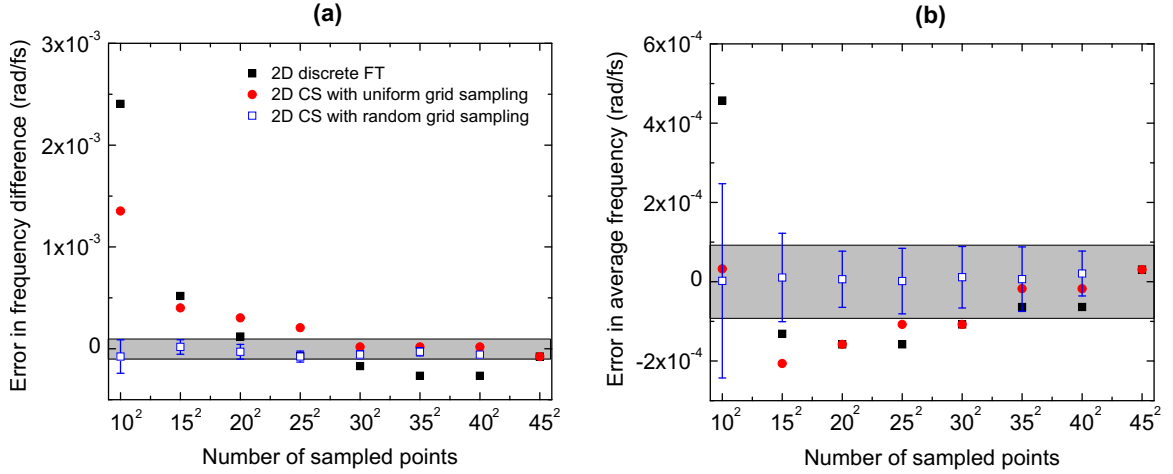
Figure 3.5: Comparison of (a) the error in the frequency difference between two optical transitions and (b) the error in the average frequency of these trans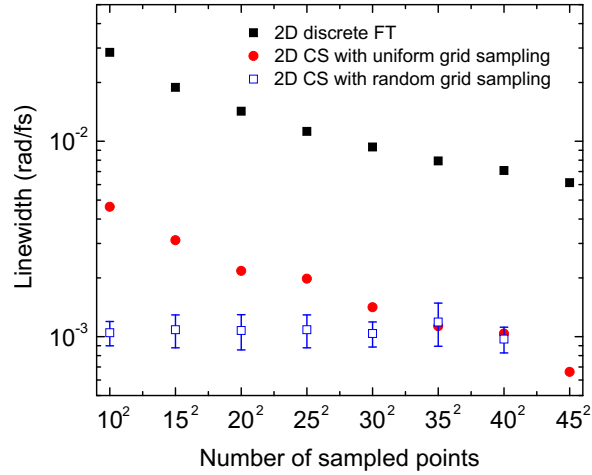itions obtained using Fourier transform and compressed sensing, with both uniform and random sampling, as the number of time-domain measurements is varied. The true values are 0.0448 rad/fs for the frequency difference and 2.3918 rad/fs for the frequency average. The discrete FT (black) and the CS with uniform grid sampling (red) are done using the first N^2 evenly-spaced points on a two-dimensional time grid with $\Delta\tau = \Delta t = 26.7$ fs. The CS with random grid sampling (blue) is done using N^2 points drawn randomly from the entire 45×45 time grid. For each random sampling data point, the frequency difference and the average frequency error are averaged over ten different random samplings; error bars denote the standard deviation. The gray area shows "exact" positions to within the precision of our frequency grid (which has spacing 0.000189 rad/fs).

45

energy difference between the highest-energy and lowest-energy peaks (similar to discrete FT). Random CS sampling, in contrast, provides the best of both worlds: a way to decrease the average sampling density without sacrificing either frequency resolution or frequency bandwidth.

Taken together, our results give a prescription for how to obtain the highest possible spectral resolution using CS. Even with our 2D CS method, the spectral resolution is still limited by the total sampling time. Though CS provides an order of magnitude improvement compared to the discrete FT, measurements collected at large τ and t will still be required to achieve high resolution. However, a major advantage of CS is that random sampling can be used to drastically cut down on the *number* of measurements needed to achieve this high resolution.

Despite the improved resolution, CS still does not resolve the natural linewidth of the peaks, which are very narrow. The issue of peak shape certainly deserves a more extensive investigation in the future, most likely by applying our 2D CS method to more complex experimental systems with intricate features that are broadened by internal structure or an environment. Another interesting continuation of our work would be to compare the CS with other techniques for resolving spectral lines, including linear prediction (LP) [107, 176, 96] and maximum entropy methods [70, 171]. Comparison to other sampling methods will require the use of many experimental signals and the development of performance metrics. This is an open research direction.

In conclusion, we have demonstrated the first application of compressed sensing to 2D optical spectroscopy experiments. Focusing on electronic transitions in an atomic ^{87}Rb vapor model system, we have shown that the 2D CS method that we have developed provides much finer resolution of the peaks in a 2D spectrum as compared to the standard discrete

FT. In particular, experimentalists can apply 2D CS as a simple "drop-in" replacement for a discrete FT to obtain an order of magnitude improvement in spectral resolution from the same samples. Moreover, CS theory and our results with quasi-random sets suggest that, in order to increase even further the amount of useful information obtained per sample, a fully random sampling strategy should be adopted for experimental measurements. We expect that 2D CS will substantially reduce the experimental effort needed to obtain well-resolved spectra, increasing the range of spectral features resolvable in ultrafast experiments, particularly closely-spaced peaks. We are confident that CS will become more widely investigated and employed in the 2D experimental ultrafast community once its easy portability and strong resolving power become more widely known.

3.2 Methods

The PM-2D FS method is described in detail in ref. [178]. A Rb vapor cell was excited by a sequence of four collinear optical pulses with adjustable inter-pulse delays (see Figure 3.1). The transmitted beam intensity was used to determine the stage positions corresponding to $\tau = t = 0$. The fluorescence emitted from the sample was detected using an avalanche photodiode. The phases of the pulse electric fields were continuously swept at distinct frequencies using acousto-optic Bragg cells, and separate reference waveforms were constructed from the resultant intensities of pulses 1 and 2, and of pulses 3 and 4, which were set to 5 kHz and 8 kHz, respectively. The reference signals were sent to a waveform mixer to construct "sum" and "difference" side-band references (13 and 3 kHz, respectively). These side-band references were used to phase-synchronously detect the fluorescence. All measurements were carried out at room temperature.

Time-resolved "sum" and "difference" signals were collected at all points on a two-

dimensional grid consisting of 45 equally-spaced coherence times τ and 45 equally-spaced detection times t (with $\Delta\tau = \Delta t = 26.7$ fs) for a series of 5 population times $T = 140$, 175, 210, 245, 280 fs. In this paper, we analyzed the "difference" signal obtained at $T = 140$ fs, though similar results were obtained for different population times as well as for "sum" signals.

For both the FT and CS calculations, we use a two-dimensional frequency grid consisting of 500 evenly-spaced points between $-2\pi/(5\Delta\tau)$ and $2\pi/(5\Delta\tau)$ for ω and 500 evenly-spaced points between $-2\pi/(5\Delta t)$ and $2\pi/(5\Delta t)$ for w. This corresponds to a frequency spacing of $\Delta\omega = \Delta w = 0.000189$ rad/fs. This frequency grid is wide enough to contain all of the peaks present in the spectrum. (For more robust numerical stability, in all CS calculations, we surrounded this grid by a coarser grid with 100 evenly-spaced points in each dimension covering the entire frequency range $[-\pi/\Delta\tau, \pi/\Delta\tau] \times [-\pi/\Delta t, \pi/\Delta t]$, excluding the part already covered by the finer grid.)

Performing the 2D CS calculations involves solving the optimization problem in equation (5.2) for different amounts of uniformly-sampled and randomly-sampled time-domain data as described in the text. Solving this optimization problem is not trivial, so we rely on the spectral projected gradient for 1-norm minimization (SPGL1) algorithm developed by van den Berg and Friedlander, which is available as free and open-source MatLab code [184]. To avoid numerical stability issues we work with a normalized BPDN problem, where the prefactor $(4\Delta\omega\Delta w/\pi^2)$ of the F matrix, equation (3.4), is left out and $\boldsymbol{h}$ is normalized. The missing factors are included in $\boldsymbol{g}$ after the solution is found.

Acknowledgments

Chapter 4

Accelerating the Computation of Bath Spectral

Densities with Super-Resolution

Apart from minor modifications, this chapter originally appeared as [115]:

Abstract

Quantum transport and other phenomena are typically modeled by coupling the system of interest to an environment, or bath, held at thermal equilibrium. Realistic bath models are at least as challenging to construct as models for the quantum systems themselves, since they must incorporate many degrees of freedom that interact with the system on a wide range of timescales. Owing to computational limitations, the environment is often modeled with simple functional forms, with a few parameters fit to experiment to yield semi-quantitative results. Growing computational resources have enabled the construction of more realistic bath models from molecular dynamics (MD) simulations. In this paper, we develop a numerical technique to construct these atomistic bath models with better accuracy and decreased cost. We apply a novel signal processing technique, known as

super-resolution, combined with a dictionary of physically-motivated bath modes to derive spectral densities from MD simulations. Our approach reduces the required simulation time and provides a more accurate spectral density than can be obtained via standard Fourier transform methods. Moreover, the spectral density is provided as a convenient closed-form expression which yields an analytic time-dependent bath kernel. Exciton dynamics of the Fenna-Matthews-Olsen light-harvesting complex are simulated with a second order time-convolutionless master equation, and spectral densities constructed via super-resolution are shown to reproduce the dynamics using only a quarter of the amount of MD data.

4.1 Introduction

Irreversible processes such as solvation, energy transfer, and chemical binding have received renewed interest in recent years. Because these processes involve large systems with many degrees of freedom, the typical approach to studying these processes is the open quantum systems formalism, in which the degrees of freedom are partitioned into a system of interest and a bath held at thermal equilibrium [28, 163]. It is commonly assumed that the system only couples weakly to the bath, making the precise nature of the bath a secondary concern in the physical theory. For example, in studying the energy transfer dynamics in a system of chromophores embedded in a protein framework, each chromophore is individually coupled to many thousands of atoms in the protein, but the system-bath formalism dramatically simplifies all of these couplings in order to make the dynamics tractable [23, 24, 166]. Renewed interest in the strong and intermediate coupling region, relevant for energy transfer in the exciton dynamics of light-harvesting complexes, has lead to various studies [87, 124, 140, 148, 36, 83, 85, 156, 1, 192, 125, 102, 167, 150, 147, 166, 135, 187, 38, 204, 151, 132, 131, 75] on the precise influence of the bath on the

51

higher systems. Higher order phonon processes, non-Markovian effects and structures in the exciton-phonon coupling change the energy transfer [49, 48, 101]. Thus, details in the bath are relevant and need to be taken into account in realistic simulations. Accordingly, our goal in this paper is to apply a recent signal-processing technique known as super-resolution to obtain realistic atomistic models of environments containing thousands of atoms at feasible computational expense. With these atomistic bath models in hand, one can begin to evaluate the importance of a realistic bath model in a physical theory.

In the approach to open quantum systems employed in this work, we model the bath by an ensemble of noninteracting harmonic oscillators. The central mathematical object of such a model is the *spectral density*, $J(\omega)$, which gives the frequency-dependent strength of system-bath coupling. The spectral density can be understood as the density of bath oscillator states at each frequency. Owing to computational limitations, most studies of open quantum systems assume an extremely simple functional form for the spectral density, such as a single broad peak covering all relevant excitonic transitions of the system. With the goal of providing more physically accurate bath models and dynamics, Valleau *et al.* has previously obtained atomistic spectral densities for the Fenna-Matthews-Olson (FMO) complex from combined Molecular Dynamics (MD) [183, 181] and time-dependent density functional theory (TDDFT) [153] simulations. However, the difficulty of this more realistic approach is the high computational cost of running expensive TDDFT calculations at every step in an MD simulation. In order to obtain a spectral density of sufficient resolution, the MD-TDDFT simulation must be run for over 40 picoseconds (ps) [113], which may become computationally intractable for larger systems.

To make progress, we first observe that a typical vibrational bath is not an arbitrary function but rather a relatively sparse collection of damped harmonic oscillators. Spar-

sity enables us to apply a novel numerical technique known as super-resolution in order to reconstruct the spectral density from much shorter MD-TDDFT simulations. Super-resolution has been applied to a broad range of scientific problems, including image [64] and video compression [136], image denoising [60], astronomy [144], microscopy [120], and medical imaging [100]. To our knowledge, this paper is the first application of super-resolution to quantum dynamics. Super-resolution provides a provably convergent algorithm for the reconstruction of signals from limited time-domain measurements using a total variation minimization procedure. Super-resolution is related to compressed sensing [112, 57, 130, 78, 145, 111, 182, 159, 122, 59, 51]. Compressed sensing is a technique designed to recover sparse signals from randomly-sampled data by minimizing the $\mathcal{L}_1$ norm of an underdetermined system of linear equations. Compressed sensing works by finding the sparsest signals consistent with the underdetermined system of equations. This usually involves an optimization problem. Despite its success in many applications, the $\mathcal{L}_1$-norm minimization of compressed sensing can result in spurious signals as it emphasizes the sparsity of the solution only. Super-resolution is a numerical method that shares the spirit of compressed sensing. The difference between superresolution and compressed sensing stems from both the choice of objective function and sampling technique. It was developed to recover sparse signals from nonrandomly undersampled data. By minimizing the $\mathcal{L}_1$-norm of the gradient of the function in addition to the $\mathcal{L}_1$ norm of the function itself, super-resolution allows for smoother solutions to the sampling problem [32, 26, 27].

Because of the ample experimental and theoretical data to compare against [183, 125, 188, 131, 2, 47, 199, 150, 164, 135, 187, 140, 166, 98, 148, 167, 102, 101, 204, 36, 156, 83, 86, 84], we apply super-resolution to the FMO light-harvesting complex of *C. tepidium* but emphasize that this technique is broadly applicable. While this paper focuses on a

vibrational bath which perturbs the energies of molecular electronic states, the techniques we introduce are generic for any model of a bath which is based on time-correlation functions.

4.2   Super-resolution of Spectral Densities

In this section, we briefly review the procedure for simulating the dynamics of open quantum systems and computing spectral densities from combined MD-TDDFT simulations. We then apply the theory of super-resolution to accelerate and improve the accuracy of these computations. Computing spectral densities from atomistic calculations, rather than from semi-empirical functional forms, enables the inclusion of molecular vibrations and other physical effects (such as solvation effects) to produce a more realistic bath model [183]. Super-resolution, in turn, brings the construction of these atomistic bath models into the realm of computational feasibility.

Armed with our more realistic bath model, we will employ a second-order time-convolution-less master equation (TCL-2) to simulate the dynamics of FMO monomer, allowing us to evaluate the physical impact of different approximations to the spectral density. TCL-2 includes non-Markovian effects up to second order in the system-bath coupling. By comparing TCL-2 with exact methods like the hierarchical equations of motion (HEOM) [177] we show that most of the relevant effects of the structured spectral density of the FMO complex are captured by TCL-2. Here we use TCL-2, since it is numerically more treatable than HEOM, in particular for structured spectral densities where HEOM becomes cumbersome and requires a high performance GPU implementation [102, 101]. We employ the

equation of motion [28, 147, 148, 164, 124, 204, 139, 180, 4, 76, 160, 168]:

$$\frac{d\rho_I(t)}{dt} = -\frac{i}{\hbar}[H_I, \rho_I] \tag{4.1}$$
$$-\frac{1}{\hbar^2}\sum_n \int_0^t d\tau \, D_n(t-\tau)[H_{In}(t), [H_{In}(\tau), \rho_I(t)]]$$

$$D_n(t) = \tag{4.2}$$
$$\int_0^\infty d\omega \, J_n(\omega)\left[\coth\left(\frac{\hbar\omega\beta}{2}\right)\cos(\omega t) - i\sin(\omega t)\right]$$

where H is the system Hamiltonian, ρ is the system density matrix, D is our bath kernel, the subscript I indicates that we are in the interaction picture, the summation runs over all sites, and $J(\omega)$ is the spectral density computed via super-resolution [98, 86, 94, 102]. The bath kernel is heavily dependent on our spectral density, causing it to play a central role in our dynamics. Therefore, a more physical bath picture should provide more physically intuitive dynamics.

In our atomistic bath model, molecular vibrations in the environment (e.g. a protein framework or solvation effects) create fluctuations in the energy gaps between the ground and excited states of the system (e.g. a set of chromophores). These time-dependent energy gaps are computed from TDDFT calculations run on each of the chromophores at each step of the MD simulation. The key object in the computation of spectral densities is the correlation function of the energy gap time series,

$$C(t) = Tr_b[\Delta(t)\Delta(0)\rho_b], \tag{4.3}$$

where $\Delta(t)$ is the time-dependent energy gap between the ground and the first excited state of the system (as calculated with TDDFT), ρ_b is the density matrix of the bath at thermal equilibrium, and $C(t)$ is the correlation function obtained after tracing over all the modes

of the bath. We discretize this equation by using an unbiased autocorrelation function,

$$C_k = \frac{1}{N-k} \sum_{i=1}^{N-k} (\Delta_i - \bar{\Delta})(\Delta_{i+k} - \bar{\Delta}), \tag{4.4}$$

where $\bar{\Delta}$ is the mean energy gap and i and k denote discrete time indices. Note that C_k involves comparing energy gaps that are k time steps apart (Δ_i and Δ_{i+k}), and $N - k$ is the total number of included comparisons.

The frequency-dependent spectral density, $J(\omega)$, is typically obtained by computing the Fourier transform of the correlation function [183]. From the definition of C_k above, it is easy to check that the correlation function is real and symmetric ($C_k = C_{N-k}$), which implies that the Fourier transform should be real and symmetric as well. Because quantum mechanical spectral densities must instead be antisymmetric and obey detailed balance, it is necessary to introduce a prefactor that enforces these two properties. Many choices are possible [22], but Valleau *et al.* have previously shown that a harmonic prefactor, $\beta\hbar\omega/2$, produces the most physical temperature dependence [183]. With this choice, the spectral density becomes the cosine transform

$$J(\omega) = \frac{\beta\hbar\omega}{2} \int_{-\infty}^{\infty} \cos(\omega t) C(t) \mathrm{dt}, \tag{4.5}$$

which characterizes the frequency-dependent coupling strength of the system to all of the nuclear vibrational modes.

The standard approach to performing this integral is the fast Fourier transform. Unfortunately, the fast Fourier transform requires sampling on a uniform grid at the Shannon sampling rate. This means that a relatively long time series, $C(t)$, must be computed in order to obtain good resolution of the spectral density in the frequency domain [34, 113]. Given the computational cost of MD simulations, and the even greater expense of running TDDFT calculations on top of these simulations, any method which can reduce the

required length of the time series $C(t)$ unplugs the computational bottleneck in deriving physically-accurate atomistic spectral densities. That is our main goal in this paper.

While reducing the amount of time required to reproduce $J(\omega)$ we also choose a basis of functions which has a convenient physical form. When decomposed into a basis of damped cosines,

$$g_{ij}(t) = e^{-\gamma_i t} \cos(\Omega_j t), \tag{4.6}$$

the function $C(t)$ is smooth and sparse. This allows for the use of the machinery of super-resolution.

To apply the super-resolution method, we discretize in time and cast our task as an inversion problem

$$C_k = \lambda_{ij} e^{-\gamma_i t} \cos(\Omega_j t_k), \tag{4.7}$$

where we seek the basis expansion coefficients λ_{ij} and have assumed Einstein summation convention over repeated indices. This can be rewritten as

$$C_k = A_{ijk} \lambda_{ij}, \tag{4.8}$$

where

$$A_{ijk} = e^{-\gamma_i t} cos(\Omega_j t_k) \tag{4.9}$$

is a matrix of damped cosines, and λ_{ij} is the set of basis coefficients we seek to recover.

The central idea of super-resolution is that the sparsity of λ_{ij} enables its full recovery even when the system $C_k = A_{ijk} \lambda_{ij}$ is underdetermined, which is to say the number of time samples C_k is significantly smaller than the number of total expansion coefficients λ_{ij} we seek to recover. Hence, we can recover the expansion coefficients on a dense grid of frequencies Ω_j and damping coefficients γ_i from fewer time samples C_k. Of the many possible solutions to our underdetermined system, super-resolution simply selects a balance

between the smoothest and sparsest (with an emphasis on smoothness) set of basis expansion

coefficients. Formally, this is done by finding the vector λ_{ij} that minimizes

$$\underset{\lambda_{ij}}{\operatorname{argmin}} \{||\nabla \lambda_{ij}||_1 + \mu||\lambda_{ij}||_1\}$$

$$\text{subject to } ||A_{ijk}\lambda_{ij} - C_k||_2 < \eta,$$

$$(4.10)$$

where the subscript 1 represents the $\mathcal{L}_1$ norm (sum of absolute values), μ represents a

sparsity penalty, $\nabla \lambda_{ij}$ represents the total variation norm, and η represents the solution

tolerance. By minimizing $||\nabla \lambda_{ij}||_1$, or total variation term, we are enforcing smoothness

in the time domain on the reconstructed signal. This throws out the "peaky" solutions

that can appear with compressed sensing [13, 155]. The total variation norm also provides

us with a provably exact technique for recovering peak position at the expense of peak

amplitude [32], which solves one of the issues seen previously with compressed sensing [13].

Recovering the expansion coefficients λ_{ij} in this manner by solving an underdetermined

matrix inversion problem takes advantage of the natural sparsity of the problem and, as

we will see in the next section, enables the construction of a well-resolved spectral density

with far less time-domain data. Even more attractive, with the λ_{ij} coefficients in hand, it

is possible to construct an analytical representation of the spectral density by taking the

cosine transform of the basis functions $g_{ij}(t)$ and applying the appropriate prefactors:

$$J(\omega) = \frac{\lambda_{ij}}{\sqrt{\pi}} \left(\frac{\beta \hbar \omega \gamma_i}{\gamma_i^2 + (\omega - \Omega_j)^2} + \frac{\beta \hbar \omega \gamma_i}{\gamma_i^2 + (\omega + \Omega_j)^2} \right),$$

$$(4.11)$$

where the Einstein summation convention has again been assumed. This is an analyti-

cal representation of the spectral density in Drude-Lorentz form, and it explicitly provides

the oscillation frequencies which characterize the system-bath coupling. We note that the

Drude-Lorentz basis naturally provides us with a width parameter, γ, that can be un-

derstood as the lifetime of oscillations in the bath. This is seen by examining the time

dependent formula, Eq. (4.7), where this γ parameter determines the strength of damping. It is important to note that in the limit as $\gamma \to 0$, we recover the cosine basis in the time domain and a Dirac delta distribution in the frequency domain. By using this super-resolution technique in concert with the Drude-Lorentz basis, we see that we can recover a small set of peaks with physically-relevant information. Additionally, the parameters that characterize the Drude-Lorentz spectral densities can be input directly into both TCL-2 and HEOM without any additional parameter fitting or numerical integration.

4.3 Numerical Methods

We employ the proposed Drude-Lorentz super-resolution method described above and apply it to a monomer of the Fenna-Matthews-Olsen (FMO) photosynthetic energy transfer complex of the green-sulfur bacterium *C. tepidium*. The FMO monomer is a system of seven chlorophyll molecules which are excitonically coupled to each other, as well as to the vibrations of the atoms in the protein framework. It functions as a molecular excitonic wire, passing excitons from the light harvesting antenna complex to the reaction center, where a biochemical cascade is initiated.

To create spectral densities for the FMO complex, we use the MD-TDDFT results of Shim *et al.* [164]. The calculations were done in an isothermal-isobaric ensemble at 77 K using the AMBER force field [50, 43]. These calculations began with a 2 ns equilibration before performing the production computations. The production steps ran for a total of 40 picoseconds with a 2 femtosecond timestep, and the optical gap was calculated for each fragment every 4 femtoseconds using TDDFT with the BLYP [19, 121, 108] functional in the 3-21G basis set in Q-Chem [161].

To perform super-resolution numerically, we require an algorithm which minimizes
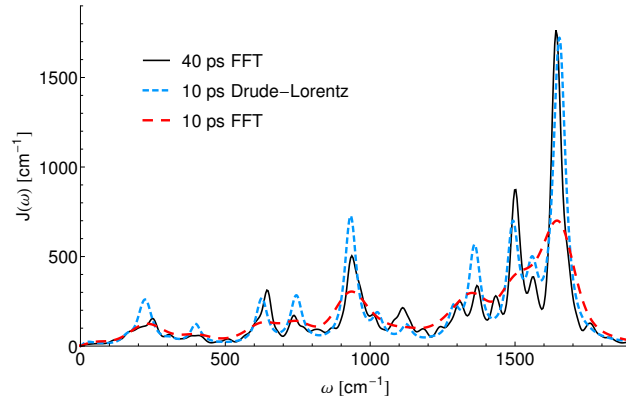
Figure 4.1: Comparison of the spectral density for site 1 of the FMO complex as a function of time and technique for spectral density recovery. Compared to the fast Fourier transform at 40 ps, much of the fine structure is easily recovered by super-resolution in the Drude-Lorentz basis, even with significant undersampling by a factor of four.
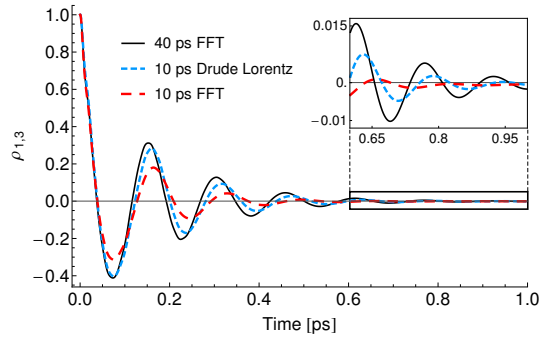
Figure 4.2: Comparison of the coherences between excitonic eigenstates 1 and 3 as a function of time and technique for spectral density recovery. Compared to the fast Fourier transform at 40 ps, the 10 ps Drude-Lorentz decomposition introduces a slight shift in oscillation frequency, but nevertheless yields more accurate dynamics than the equivalently-sampled fast Fourier transform at 10 ps.
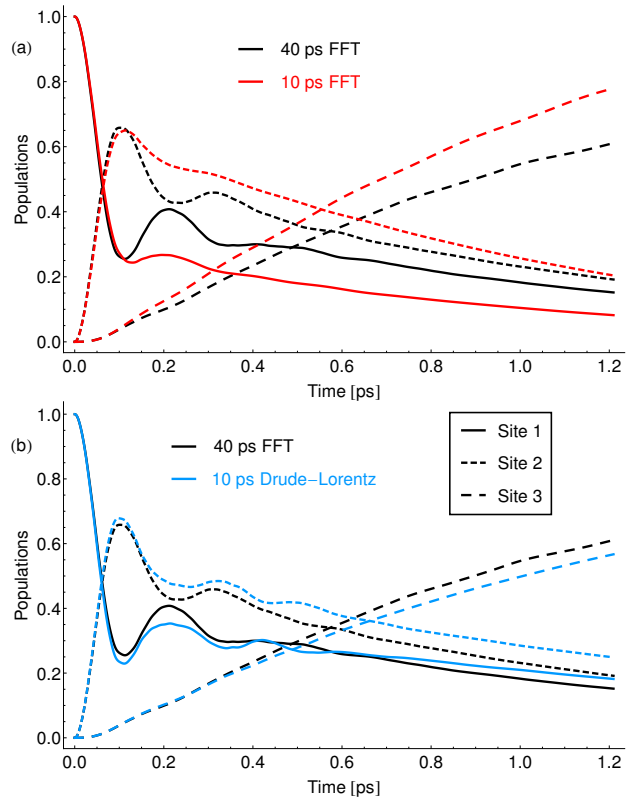
Figure 4.3: Comparison of the populations for sites 1-3 as a function of time and technique for spectral density recovery Compared to the fast Fourier transform at 40 ps, the 10 ps Drude-Lorentz decomposition recovers the overall shape and provides much more faithful dynamics than the equivalently-sampled fast Fourier transform at 10 ps.

the total variation norm to solve the minimization problem described by eq. (4.10). In our implementation, we use the two step iterative shrinkage thresholding (TwIST) algorithm [26, 27], which combines computational efficiency with strong convergence. To construct the measurement matrix A described in eq. (4.9), we must select a grid of possible frequencies ($\{\Omega_j\}$) and linewidths ($\{\gamma_i\}$). In our implementation, we use a grid of frequencies ranging from 0 to 2000 cm^{-1} in 2 cm^{-1} intervals, and a grid of linewidths ranging from 0 to 160 cm^{-1} in 6 cm^{-1} intervals. We assume that our calculations are converged when $\eta < 10^{-7}$ (in eq. (4.10)), or the solution vector remains constant for 100 iterations. Finally, we perform an $\mathcal{L}_2$ minimization of $A_{ijk}\lambda_{ij} - C_k$ while freezing the recovered nonzero basis functions, allowing us to further minimize the error. We refer to this procedure as debiasing because it partly removes the bias towards sparsity and smoothness introduced by the $\mathcal{L}_1$ minimization. This debiasing procedure reduces our solution tolerance to $\eta < 10^{-9}$, allowing convergence to a better solution. It is important to note that, in general, the super-resolution technique is robust to an over-complete basis.

4.4   Results

Fig. 4.1 shows the results of employing the Drude-Lorentz super-resolution method to recover the spectral density for site 1 of FMO. The figure compares Drude-Lorentz super-resolution with 10 ps of MD to a standard fast Fourier transform approach with both 10 and 40 ps of MD. We take the fast Fourier transform with 40 ps of MD as our standard for comparison. By comparing the two methods with 10 ps of MD, it is clear that super-resolution resolves more features of the spectral density than the standard fast Fourier transform from the same amount of time-domain data. Moreover, super-resolution captures most of the features of the fast Fourier transform with the full 40 ps of MD: we

see the expected CO stretch at 1600 cm^{-1}, which we attribute to the amides in the protein scaffold, as well as all of the other major peaks in the spectral density. We attribute a significant amount of the error in our spectral density reconstruction to the fact that the truncated MD series does not explore the phase space as thoroughly in only 10 ps.

The Drude-Lorentz basis also provides significant sparsity gains in comparison to the cosine basis: we require only 56 Drude-Lorentz peaks to create the spectral density given in Fig. 4.1 whereas. This sparsity provides a significant computational advantage for excitonic propagation in both hierarchical equations of motion (HEOM) [177] and second order time-convolutionless master equation [28](TCL-2) approaches because the propagations scale factorially and linearly, respectively, as a function of the number of peaks included. In the excitonically accessible regime of 0-540cm^{-1}, we recover only 20 Drude-Lorentz peaks, and six of them have amplitudes that are two orders of magnitude smaller than the rest. These Drude-Lorentz peaks can be entered directly into master equation simulations, including HEOM codes, without the need to perform any intermediate fitting. In summary, super-resolution yields a well-resolved spectral density using less time-domain data than is required by the standard fast Fourier transform approach and precludes the need for additional fitting.

As mentioned above, the TCL-2 propagation of the exciton dynamics of the FMO complex, with the Hamiltonian coming from [2], was carried out using the Drude-Lorenz spectral densities obtained from super-resolution. We propagated 1 ps of dynamics and obtained the populations of sites 1-3, as well as the coherence between sites 1 and 3.

Fig. 4.2 shows the coherence between excitonic eigenstates 1 and 3 as a function of time. Compared to the 40 ps fast Fourier transform, we see that the 10 ps Drude-Lorentz super-resolution more faithfully reproduces the coherence dynamics than the 10 ps fast

Fourier transform, both in terms of the oscillation frequency and the overall damping. The fast Fourier transform with 10 ps of MD data introduces serious overdamping as well as a significant shift in oscillation frequency. In contrast, the Drude-Lorentz expansion with 10 ps of MD data introduces only a small shift in oscillation frequency, resulting in more accurate coherence dynamics overall. We attribute most of the discrepancies to slight relative differences in the reorganization of each site between spectral densities constructed with 10 and 40 ps of MD data. It appears that while the oscillations are extremely sensitive to the relative reorganization energies between the sites, the damping is more dependent on the fine structure of the spectral densities. The Drude-Lorentz super-resolution (10 ps MD-TDDFT data) reproduces the coherence life-times obtained by fast Fourier transform recovered using all 40 ps of MD-TDDFT data – representing a factor of four improvement.

The contrast between the two approximation techniques becomes even more significant when we simulate dynamics beginning with an exciton fully localized on site 1. In Fig. 4.3, we have plotted the populations of the first three sites as a function of time. The Drude-Lorentz expansion with 10 ps of MD yields good qualitative agreement with our standard of comparison whereas the fast Fourier transform. The fast Fourier transform on 10 ps overestimates population transfer to site 3 at short times and grows much more quickly from there, whereas the Drude-Lorentz expansion slightly under predicts the population transfer at long times. We attribute these errors in the asymptotic behavior to slight differences in the reorganization energies for the spectral densities of each of the sites, since each site is embedded in a different enviornment, the reorganization process of the individual pigments is different. This sensitivity affects overall dissipation and even small changes in the spectral density of the Drude-Lorentz expansion (10 ps) when compared to the standard of comparison affects energy relaxation. Beyond that, the Drude-Lorentz

expansion is capable of reproducing the oscillations at 0.2 and 0.4 ps in the data for sites 1 and 2 whereas the fast Fourier transform reproduces them less faithfully. In summary, the Drude-Lorentz super-resolution technique provides us with much more physical behavior.

4.5   Conclusions

We have shown that the Drude-Lorentz super-resolution method provides significant computational advantages for the construction of atomistic bath models. In particular, the super-resolution calculations require only 10 ps of MD-TDDFT simulations to obtain reasonable atomistic spectral densities and system dynamics; this is one quarter the amount of data needed in standard fast Fourier transform-based calculations. Ultimately, this will permit the use of more physically accurate calculations or larger systems. Given the computational expense of running TDDFT calculations at every MD simulation step, we believe that the super-resolution method will enable the treatment of larger systems than previously possible.

One of the most significant advantages of our super-resolution method is the decomposition of these atomistic spectral densities into a naturally-sparse basis of Drude-Lorentz oscillators. This makes it easy to perform fast master equation simulations within either the TCL-2 or HEOM formalisms by exploiting analytic integrals of the spectral density. Beyond this, we also directly extract physically-important parameters such as the coherence lifetimes of all the oscillators in the bath. In the future, it is easy to imagine turning this technique on its head to create new spectral densities in a constructive fashion from a set of Drude-Lorentz oscillators.

Acknowledgments

Part II

Compressed Sensing for Quantum

Chemistry

Chapter 5

Compressed Sensing for the Fast Computation of Matrices: Application to Molecular Vibrations

Abstract

This article presents a new method to compute matrices from numerical simulations based on the ideas of sparse sampling and compressed sensing. The method is useful for problems where the determination of the entries of a matrix constitutes the computational bottleneck. We apply this new method to an important problem in computational chemistry: the determination of molecular vibrations from electronic structure calculations, where our results show that the overall scaling of the procedure can be improved in some cases. Moreover, our method provides a general framework for bootstrapping cheap low-accuracy calculations in order to reduce the required number of expensive high-accuracy calculations, resulting in a significant $3\times$ speed-up in actual calculations.

5.1 Introduction

Matrices are one of the most fundamental objects in the mathematical description of nature, and as such they are ubiquitous in every area of science. For example, they arise naturally in linear response theory as the first term in a multidimensional Taylor series, encoding the response of each component of the system to each component of the stimulus. Hence, in many scientific applications, matrices contain the essential information about the system being studied.

Despite their ubiquity, the calculation of matrices often requires considerable computational effort. Returning to the linear response theory example, it might be necessary to individually calculate the response of every component of the system to every component of the stimulus and, depending on the area of application, each individual computation may itself be quite expensive. The overall expense stems from the fact that evaluating a matrix of dimension $N \times M$ requires, in principle, the individual evaluation of $N \times M$ elements. But this does not always have to be the case.

For example, if we know *a priori* the eigenvectors of a $N \times N$ diagonalizable matrix, then we can obtain the full matrix by only calculating the N diagonal elements. Similarly, a sparse matrix, which contains many zero elements, can be evaluated by calculating only the non-zero elements, if we know in advance where such elements are located. In this article, we present a general approach that can produce a considerable reduction in the cost of constructing a matrix in many scientific applications by substantially reducing the number of elements that need to be calculated.

The key numerical procedure of our approach is a method to cheaply recover sparse matrices with a cost that is essentially proportional to the number of non-zero elements. The matrix reconstruction procedure is based on the increasingly popular compressed sens-

ing approach [34, 57, 35, 46], a state-of-the-art signal processing technique developed to minimize the amount of data that needs to be measured to reconstruct a sparse signal.

Although the theory of compressed sensing is extensive and well-developed, the use of compressed sensing and sparse sampling methods in scientific development have been dominated by experimental applications, including multidimensional nuclear magnetic resonance [92, 81], super-resolution microscopy [205], and other applications in spectroscopy and beyond [82, 56, 69, 159, 155, 169, 14, 193]. However compressed sensing is also becoming a tool for computational applications [13, 7, 157, 128, 115, 63, 15]. In particular, in previous work we have shown that compressed sensing can also be used to reduce the amount of computation in numerical simulations [13].

In this article, we apply compressed sensing to the problem of computing matrices. This method has two key properties. First, the cost of the procedure is quasi-linear with the size of the number of non-zero elements in the matrix, without the need to know *a priori* the location of the non-zero elements. Second, the reconstruction is exact. Furthermore, the utility of the method extends beyond the computation of *a priori* sparse matrices. In particular, the method suggests a new computing paradigm in which one develops methods to find a basis in which the matrix is known or suspected to be sparse, based on the characteristics and prior knowledge of the matrix, and then afterwards attempts to recover the matrix at lower cost.

To demonstrate the power of our approach, we apply these ideas to an important problem in quantum chemistry: the determination of the vibrational modes of molecules from electronic structure methods. These methods require the calculation of the matrix of the second derivatives of the energy with respect to the nuclear displacements, known as the force-constant or Hessian matrix. This matrix is routinely obtained in numerical simulations

70

by chemists and physicists, but it is relatively expensive to compute when accurate quantum mechanical methods are used.

The search for more efficient methods of computing Hessian matrices has a long history. One of the earliest advances was the development of theoretical methods to compute the derivatives of the matrix analytically, rather than numerically, first in one of the simplest quantum mechanical methods known as Hartree-Fock theory [66, 141], then in more sophisticated higher-accuracy correlated methods [72, 73, 103], and later in modern density functional theory [74]. These analytical derivative techniques have been further optimized by systematically exploiting molecular symmetries [174] and by organizing calculations more efficiently by working directly in the so-called atomic orbital basis [134, 65]. More recent developments have applied Davidson methods to compute only those vibrational modes relevant to searching for the transition state in a chemical reaction while sidestepping the computation of a full Hessian matrix [149, 162]. Finally, another research frontier involves developing methods that scale favorably with system size for computing energies [106] and response properties such as molecular vibrations [21, 133], with the ultimate goal of achieving practical techniques that exhibit linear scaling.

Our approach exploits the sparsity of the Hessian matrix and and cheap auxiliary calculations to further improve the efficiency of computing the vibrational modes of molecules; moreover, our approach is compatible with and complementary to some of the aforementioned techniques. At the same time, our method provides a general framework for bootstrapping cheap low-accuracy calculations to reduce the required number of expensive high-accuracy calculations, something which previously was not possible to do in general.

We begin by discussing how compressed sensing makes it practical to take a new approach for the calculation of matrices based on finding strategies to make the matrix sparse.

Next, we introduce the mathematical foundations of the method of compressed sensing and apply them to the problem of sparse matrix reconstruction. This is the numerical tool that forms the foundation of our approach. Finally, we illustrate these new ideas by applying them to the problem of obtaining molecular vibrations from quantum mechanical simulations.

5.2 Finding a Sparse Description of the Problem

The first step in our approach is to find a representation for the problem where the matrix to be calculated is expected to be sparse. In general, finding this *sparsifying basis* is specific to each problem and ranges from trivial to quite complex; it has to do with the knowledge we have about the problem or what we expect about its solution.

Leveraging additional information about a problem is an essential concept in compressed sensing, but it is also a concept that is routinely exploited in numerical simulations. For example, in quantum chemistry it is customary to represent the orbitals of a molecule in a basis formed by the orbitals of the atoms in the molecule [173], which allows for an efficient and compact representation and a controlled discretization error. This choice comes from the notion that the electronic structure of the molecule is roughly described by "patching together" the electronic structure of the constituent atoms.

An ideal basis in which to reconstruct a matrix is the basis of its eigenvectors, or *eigenbasis*, as this basis only requires the evaluation of the diagonal elements to obtain the entire matrix. Of course, finding the eigenbasis requires knowing the matrix in the first place, so reconstructing a matrix in its eigenbasis is not practically useful. However, in many cases it is possible to obtain reasonable approximations to the eigenvectors (an idea which also forms the basis of perturbation theory in quantum mechanics). The approximate

eigenbasis probably constitutes a good sparsifying basis for many problems, as we expect the matrix to be diagonally dominant, with a large fraction of the off-diagonal elements equal to zero or at least small.

Since the determination of an approximate eigenbasis depends on the specific problem at hand, a general prescription is difficult to give. Nevertheless, a few general ideas could work in many situations. For example, in iterative or propagative simulations, results from previous iterations or steps could be used to generate a guess for the next step. Alternatively, cheap low-accuracy methods can be used to generate a guess for an approximate eigenbasis. In this case, the procedure we propose provides a framework for bootstrapping the results of a low-cost calculation in order to reduce the required number of costly high-accuracy calculations. This last strategy is the one we apply to the case of molecular vibrations.

What makes looking for sparsifying basis attractive, even at some computational cost and code-complexity overhead, are the properties of the recovery method. First, the cost of recovering the matrix is roughly proportional to its sparsity. Second, the reconstruction of the matrix is always exact up to a desired precision; even if the sparsifying basis is not a good one, we eventually converge to the correct result. The penalty for a bad sparsifying basis is additional computation, which in the worst case makes the calculation as costly as if compressed sensing were not used at all. This feature implies that the method will almost certainly offer some performance gain.

There is one important qualification to this gain. For some matrices, there is a preferred basis in which the matrix is cheaper to compute, and the extra cost of computing its elements in a different basis might offset the reduction in cost offered by compressed sensing.

5.3 Compressed Sensing for Sparse Matrices

Once a sparse representation for the matrix is known, the numerical core of our method for the fast computation of matrices is the application of compressed sensing to calculate sparse matrices without knowing *a priori* where the non-zero elements are located. Related work has been presented in the field of compressive principal component pursuit [189, 33, 203, 191], which focuses on reconstructing matrices that are the sum of a low-rank component and a sparse component. Our work instead outlines a general procedure for reconstructing any sparse matrix by measuring it in a different basis.

Suppose we wish to recover a $N \times N$ matrix A known to be sparse in a particular orthonormal basis $\{\psi_i\}$ (for simplicity we restrict ourselves to square matrices and orthonormal bases). Without any prior knowledge of where the S non-zero elements of A are located, it might appear that we need to calculate all N^2 elements, but this is not the case.

In a different orthonormal basis $\{\phi_i\}$, the matrix A has a second representation B given by

$$B = PAP^T \ , \tag{5.1}$$

where P is the orthogonal change-of-basis matrix from the basis $\{\psi_i\}$ to the basis $\{\phi_i\}$. Note that in general B is not sparse.

If we regard A and B as N^2-element *vectors*, it is easy to see that the change-of-basis transformation from A to B given by equation (5.1) is linear. This fact enables us to use the machinery of compressed sensing to reconstruct the full matrix A by measuring only *some* of the entries of B.

For a fully general matrix A, we would indeed need to measure all the entries of B so that A could be recovered by inverting the linear system in equation (5.1). However, because the matrix A is sparse, compressed sensing lets us to do better. The idea is to measure

74

only *some* of the entries of B, which means that equation (5.1) is now underdetermined and admits many possible solutions for A. From these many solutions for A, we simply select the *sparsest* one; it has been proven in general [34, 57, 35, 46] that this can be done in a highly computationally efficient way by minimizing the sum of the absolute values of the entries of A. The key insight of compressed sensing is that, as more entries of B are measured, the sparsest matrix A which satisfies the underdetermined system converges to the true solution of the full system. Moreover, this convergence typically happens long before the matrix B has been fully sampled, particularly if the true matrix A is quite sparse.

The compressed sensing reconstruction is done by solving the so-called basis pursuit (BP) problem [46, 184],

$$\min_A ||A||_1 \quad \text{subject to} \quad (PAP^T)_{ij} = B_{ij} \quad \forall \, i, j \in W \, , \tag{5.2}$$

where the 1-norm is considered as a *vector* norm ($||A||_1 = \sum_{i,j} |A_{ij}|$), the change-of-basis matrix P is known, and W is a set of randomly measured entries in matrix B which are known. We assume that we have some method of computing the entries of B, but that the method is expensive, and we would therefore like recover A while computing as few of them as possible.

The size of the set W, a number that we call M, is the number of matrix elements of B that are sampled. M determines the quality of the reconstruction of A. From compressed sensing theory we can find a lower bound for M as a function of the sparsity of A and the change-of-basis transformation.

One important requirement for compressed sensing is that the sparse basis $\{\psi_i\}$ for A and the measurement basis $\{\phi_i\}$ for B should be *incoherent*, meaning that the maximum

Sparse matrix A to recover

Incoherent basis
$B = PAP^{\mathrm{T}}$

Apply
compressed
sensing

Undersample matrix B
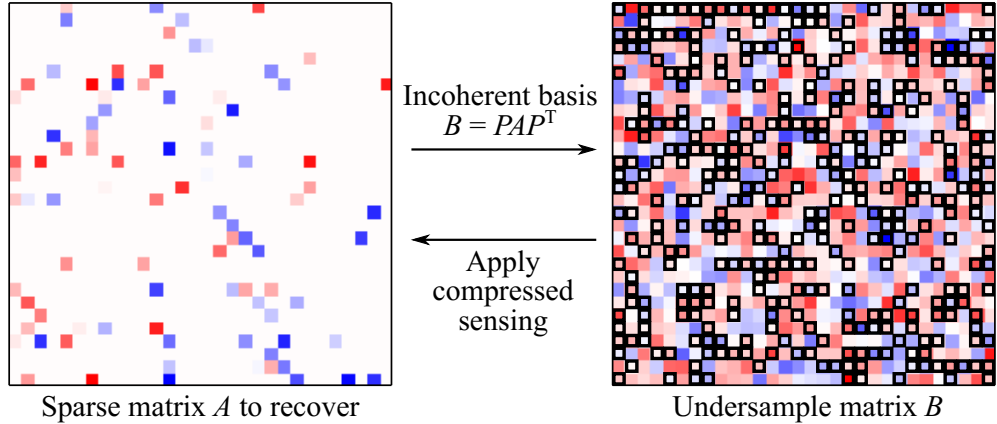
Figure 5.1: General scheme for the recovery of a sparse matrix A via compressed sensing. Rather than sampling A directly, the key is to sample the matrix B which corresponds to A expressed in an different (known) basis. Recovery of A from the undersampled entries of B proceeds via compressed sensing by solving equation (5.2).

overlap between any vector in $\{\psi_i\}$ and any vector in $\{\phi_i\}$

$$\mu = \sqrt{N} \max_{i,j} \langle \psi_i | \phi_j \rangle \tag{5.3}$$

should be as *small* as possible (in general μ ranges from 1 to $\sqrt{N}$). Intuitively, this incoherence condition means that the change-of-basis matrix P should thoroughly scramble the entries of A to generate B.

It can be proven [35] that the number of entries of B which must be measured in order to fully recover A by solving the BP problem in equation (5.2) scales as

$$M \propto \mu^2 S \log N^2 \ . \tag{5.4}$$

This scaling equation encapsulates the important aspect of compressed sensing: if a proper measurement basis is chosen, the number of entries which must be measured scales linearly with the sparsity of the matrix and only depends weakly on the full size of the matrix. For the remainder of this paper, we always choose our measurement basis vectors to be the

discrete cosine transform (DCT) of the sparse basis vectors, for which the parameter μ is equal to $\sqrt{2}$. The DCT is a common transformation chosen for compressed sensing because it is easy to implement, with fast and readily-available algorithms for its computation, and because it guarantees that the sparse basis and the measurement basis are highly incoherent. Intuitively, this ensures that each element from the measurement matrix B contains as much information as possible about the elements of the sparse matrix A. A small value of μ such as $\sqrt{2}$ cuts down on the number of entries of B which must be measured in order to fully recover the sparse matrix A. Of course, our method is independent of this choice and could work with other basis transformations as well.

In order to study the numerical properties of the reconstruction method we performed a series of numerical experiments. We generate 100×100 matrices of varying sparsity with random values drawn uniformly from the interval $[-1, 1]$ and placed in random locations in the matrix. Matrix elements were then sampled in the DCT measurement basis, and an attempt was made to recover the original sparse matrix by solving the basis pursuit problem in equation (5.2).

Figure 5.2 illustrates the percent of matrix elements had to be sampled for accurate recovery of the sparse matrix compared with other recovery approaches. If no prior knowledge of a matrix is used for its recovery, then one simply measures each entry; this is the current paradigm in many scientific applications. If one knows exactly where the non-zeros in a sparse matrix are located, one can simply measure those elements. Compressed sensing interpolates between these two extremes: it provides a method for recovering a sparse matrix when the locations of the non-zeros are not known in advance. Although this lack of knowledge comes with a cost, the recovery is still considerably cheaper than measuring the entire matrix.
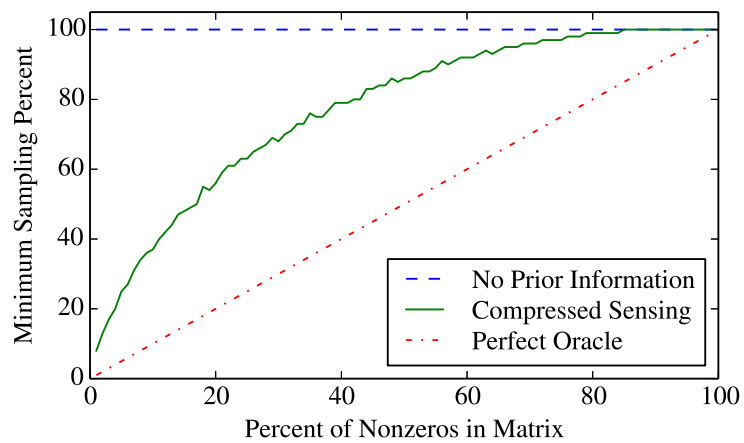
Figure 5.2: Percent of entries that must be sampled for accurate recovery of a matrix as a function of sparsity. Comparison between compressed sensing and two limiting cases: "no prior knowledge" of sparsity and the "perfect oracle" who reveals where all non-zero entries are located. Each point on the compressed sensing curve is an average of ten different randomizations. The accuracy criterion is a relative error in the Frobenius norm smaller than 10^{-7}.

5.4 Application: Molecular Vibrations

Calculating the vibrations of a molecule, both the frequencies and normal modes, is one of the most ubiquitous tasks in computational chemistry [190]. Integrated into nearly all computational chemistry packages, including the Q-Chem package used for this study [161], molecular vibrations are computed by theoretical and experimental chemists alike. Chemists routinely optimize molecular geometries to find minimal energy conformations; computing and confirming the positivity of all vibrational frequencies is the standard method of assuring that a local minimum has been found. Another common task is to find the transition state for a proposed reaction: here it is also necessary to compute the vibrations to find one mode with an imaginary frequency, confirming the existence of a local maximum along the reaction coordinate [52]. Despite the centrality of molecular vibrations in computational chemistry, it remains one of the most expensive computations routinely performed by chemists.

The core of the technique lies in calculating the matrix of the mass-weighted second derivatives of the energy with respect to the atomic positions

$$H_{Ai,Bj} = \frac{1}{\sqrt{M_A M_B}} \frac{\partial^2 E(\boldsymbol{R}^1, \ldots, \boldsymbol{R}^N)}{\partial R_i^A \partial R_j^B} \tag{5.5}$$

where $E(\boldsymbol{R}^1, \ldots, \boldsymbol{R}^N)$ is the ground-state energy of the molecule, R_i^A is coordinate i of atom A, and M_A is its mass. Hence, the Hessian is a real $3N \times 3N$ matrix where N is the number of atoms in the molecule. When the molecule is in a minimum energy conformation, the eigenvectors of the Hessian correspond to the vibrational modes of the molecule, and the square root of the eigenvalues correspond to the vibrational frequencies [52].

Our goal, therefore, is to understand how our approach can reduce the cost of computing the Hessian matrix of a molecule. We achieve this understanding in two complementary

ways. First, for a moderately-sized molecule, we outline and perform the entire numerical procedure to show in practice what kinds of speed-ups may be obtained. Second, for large systems, we investigate the ability of compressed sensing to improve how the cost of computing the Hessian scales with the number of atoms.

Calculating the Hessian requires a method for obtaining the energy of a given nuclear configuration. There exist many methods to chose from, which offer a trade-off between accuracy and computational cost. Molecular mechanics approaches, which model the interactions between atoms via empirical potentials [52], are computationally cheap for systems of hundreds or thousands of atoms, while more accurate and expensive methods explicitly model the electronic degrees of freedom at some level of approximated quantum mechanics, such as methods based on density functional theory (DFT) [80, 97, 20] or wavefunction methods [173]. We focus on these quantum mechanical approaches, since there the computation time is dominated by the calculation of the elements of the Hessian matrix, making it an ideal application for our matrix-recovery method.

To recover a quantum mechanical Hessian efficiently with compressed sensing, we need to find a basis in which the matrix is sparse. While we might expect to the Hessian to have some degree of sparsity in the space of atomic Cartesian coordinates, especially for large molecules, we have found that it is possible to find a better basis. The approach we take is to use a basis of approximated eigenvectors generated by a molecular mechanics computation, employing the common MM3 force field [6], which provides a cheap approximation to the eigenvectors of the quantum mechanical Hessian [158]. This is illustrated in Figure 5.3 for the benzene molecule (C_6H_6). The figure compares the quantum mechanical Hessian in the basis of atomic Cartesian coordinates with the same matrix in the approximate eigenbasis obtained via an auxiliary molecular mechanics computation. The matrix in the molecular
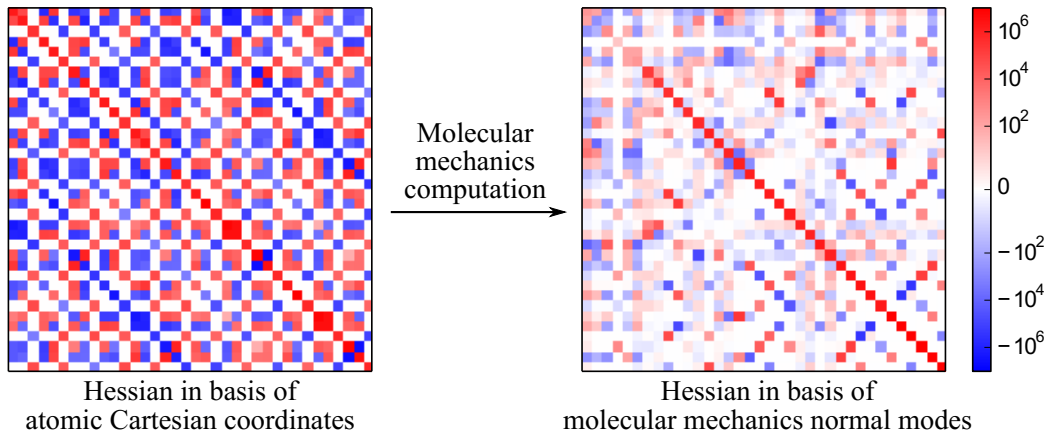
Figure 5.3: The quantum mechanical Hessian of benzene in the basis of atomic Cartesian coordinates (on the left) and in the basis of molecular mechanics normal modes (on the right). Since the molecular mechanics normal modes form approximate eigenvectors to the true quantum mechanical Hessian, the matrix on the right is sparse (close to diagonal) and therefore well-suited to recovery via compressed sensing.

mechanics basis is much sparser, and is therefore better suited to recovery via compressed sensing.

The second derivatives of the energy required for the Hessian, equation (5.5), can be calculated either via finite differences, generating what are known as *numerical* derivatives, or using perturbation theory, generating so-called *analytical* derivatives [142, 89, 18, 90]. A property of the calculations of the energy derivatives is that the numerical cost does not depend on the direction they are calculated. This can be readily seen in the case of finite differences, as the cost of calculating $E(\boldsymbol{R}^1, \ldots, \boldsymbol{R}^j + \Delta^j, \ldots, \boldsymbol{R}^N)$ is essentially the same as computing $E(\boldsymbol{R}^1 + \Delta^1, \ldots, \boldsymbol{R}^j + \Delta^j, \ldots, \boldsymbol{R}^N + \Delta^N)$. As discussed previously, this ability to compute matrix elements at a comparable cost in any desired basis is an essential requirement of our method.

A second property of both numerical and analytical derivatives that appears in vari-

ational quantum chemistry formalisms like DFT or Hartree-Fock is that each calculation yields a full column of the Hessian, rather than a single matrix element. Again, this is easy to see in finite difference computations. We can write the second derivative of the energy as a first derivative of the force

$$\frac{\partial^2 E(\boldsymbol{R}^1,\ldots,\boldsymbol{R}^N)}{\partial R_i^A \partial R_j^B} = -\frac{\partial F_j^B(\boldsymbol{R}^1,\ldots,\boldsymbol{R}^N)}{\partial R_i^A} \ . \tag{5.6}$$

By the Hellman-Feynman theorem [77, 62], and the appropriate correction for the Pulay forces [142], a single energy calculation yields the forces acting over *all* atoms, so the evaluation of equation (5.6) by finite differences for fixed A and i yields the derivatives for all values of B and j, a whole column of the Hessian. An equivalent result holds for analytic derivatives obtained via perturbation theory [18, 90]. Thus, our compressed sensing procedure for this particular application focuses on measuring random columns of the quantum mechanical Hessian rather than individual random entries.

The full compressed sensing procedure applied to the calculation of a quantum mechanical Hessian is implemented as follows:

1. Calculate approximate vibrational modes using molecular mechanics.

2. Transform the approximate modes using the DCT matrix.

3. Randomly select a few of the transformed modes.

4. Calculate energy second derivatives along these random modes to yield random columns of the quantum mechanical Hessian.

5. Apply compressed sensing to rebuild the full quantum mechanical Hessian in the basis of approximate vibrational modes.

6. Transform the full quantum mechanical Hessian back into the atomic coordinate basis.

7. Diagonalize the quantum mechanical Hessian to obtain the vibrational modes and frequencies.

The optimal number of random modes can be selected iteratively, repeating steps 3–7 adding more random modes each time until convergence is reached. See the Supporting Information for details.

Figure 5.4 illustrates the results of applying our Hessian recovery procedure to anthracene ($C_{14}H_{10}$), a moderately-sized polyacene consisting of three linearly fused benzene rings. The top panel illustrates the vibrational frequencies obtained by the compressed sensing procedure outlined above for different extents of undersampling of the true quantum mechanical Hessian. Even sampling only 25% of the columns yields vibrational frequencies that are close to the true quantum mechanical frequencies, and much closer than the molecular mechanics frequencies. The bottom-left panel illustrates the error in the vibrational frequencies from the true quantum mechanical frequencies. Sampling only 30% of the columns gives a maximum frequency error of less than 3 cm^{-1}, and sampling 35% of the columns yields nearly exact recovery. The bottom-right panel illustrates the error in the normal modes. Once again, sampling only 30% of the columns gives accurate recovery of all vibrational normal modes to within 1%. In short, our compressed sensing procedure applied to anthracene reduces the number of expensive quantum mechanical computations by a factor of three. The additional cost of the molecular mechanics computation and the compressed sensing procedure, which take a few seconds, is negligible compared to the reduction in cost for the computation of the Hessian which for anthracene takes on the order of hours.

Having shown that our compressed sensing procedure gives a $3\times$ speed-up for a moderately-sized organic molecule, we now move to larger systems and investigate how the cost of com-
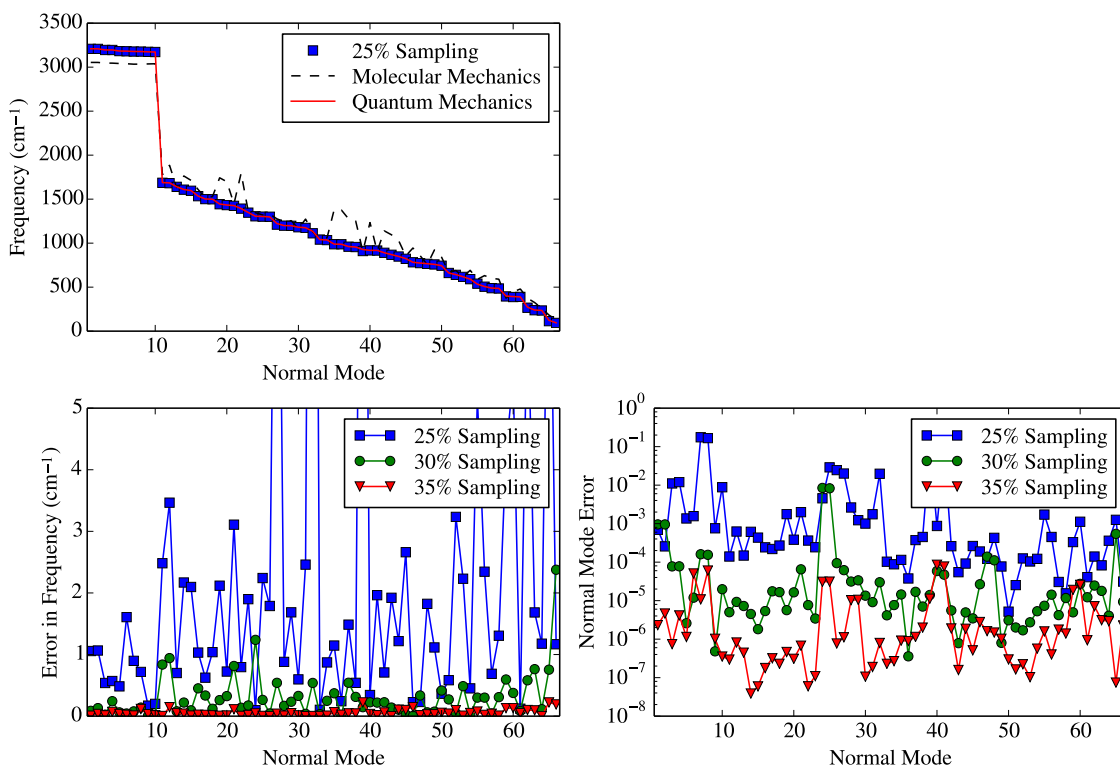
Figure 5.4: Results of applying our compressed sensing procedure to the vibrational modes and frequencies of anthracene. (Top) Even by sampling only 25% of the quantum mechanical Hessian, the vibrational frequencies obtained via compressed sensing converge to those of the true quantum mechanical Hessian. (Bottom Left) Error in vibrational frequencies for different extents of undersampling. When only 30% of the columns are sampled, the maximum error in frequency is within 3 cm^{-1}, and with 35% sampling, the recovery is essentially exact. (Bottom Right) Error in vibrational normal modes for different extents of undersampling on a logarithmic scale; the error is calculated as one minus the overlap (dot product) between the exact quantum mechanical normal mode and the normal mode obtained via compressed sensing. Once 30% of the columns are sampled, the normal modes are recovered to within 1% accuracy.

puting the Hessian scales with the number of atoms. In the absence of compressed sensing, if the entries of the Hessian must be calculated independently, the cost of calculating the Hessian would scale as $O(N^2) \times OE$, where OE is the cost of computing the energy of a given nuclear configuration (the cost of analytical and numerical derivatives usually have the same scaling). For example, for a DFT-based calculation, OE is typically $O(N^3)$. However, since many quantum mechanical methods obtain the Hessian one column at a time, only $O(N)$ calculations are required, so the scaling is improved to $O(N) \times OE$.

How does compressed sensing alter this scaling? From equation (5.4), the number of matrix elements needed to recover the Hessian via compressed sensing scales as $O(S \log N)$, where S is number of non-zero elements in the Hessian, so the net scaling is $O(S \log N) \times OE$. By obtaining the Hessian one column at a time, we expect the net scaling to improve to $O(S/N \log N) \times OE$. However, we should note that equation (5.4) is only valid in principle for the random sampling of elements, and it is not necessarily valid for a random column sampling. This scaling result illustrates the critical importance of recovering the Hessian in a sparse basis, with S as small as possible. So what is the smallest S that can reasonably be achieved?

For many large systems, the Hessian is already sparse in the basis of atomic Cartesian coordinates. Since the elements of the Hessian are *partial* second derivatives of the energy with respect to the positions of two atoms, only direct interactions between the two atoms, with the positions of all other atoms held fixed, must be taken into account. For most systems we expect that this direct interaction has a finite range or decays strongly with distance. Note that this does not preclude collective vibrational modes, which can emerge as a result of "chaining together" direct interactions between nearby atoms.

If we assume that a system has a finite range interaction between atoms, and since each

atom has an approximately constant number of neighbors, irrespective of the total number of atoms in the molecule, the number of non-zero elements in a single column of the Hessian should be constant. Hence, for large molecules, the sparsity S of the Hessian would scale linearly with the number of atoms N. Putting this result into $O(S/N \log N) \times OE$ yields a best-case scaling of $O(\log N) \times OE$, which is a significant improvement over the original $O(N) \times OE$ in the absence of compressed sensing.

To study the validity of our scaling results we have performed numerical calculations on a series of polyacene molecules, which are aromatic compounds made of linearly fused benzene rings. For polyacenes ranging from 1 to 15 rings, Figure 5.5 illustrates the average number of non-zeros per column in the Hessian matrices obtained via molecular mechanics and quantum mechanical calculations in the basis of atomic coordinates. In the molecular mechanics Hessians, the average sparsity per column approaches a constant value as the size of the polyacene increases, consistent with each atom having direct interaction with a constant number of other atoms.

Since the molecular mechanics Hessians illustrate the best-case scenario in which the sparsity S scales linearly with the number of atoms N, we attempted to recover these Hessians directly in the basis of atomic coordinates via the compressed sensing procedure we have outlined by sampling columns in the DCT basis. Figure 5.6 illustrates the number of columns which must be sampled to recover the Hessians to within a relative error of 10^{-3} as a function of the size of the polyacene. Far fewer than the total number of columns in the entire matrix need to be sampled. Even more attractive is the fact that the number of columns grows quite slowly with the size of the polyacene, consistent with the best-case $O(\log N) \times OE$ scaling result obtained above. This result indicates that our compressed sensing approach is especially promising for the calculation of Hessian matrices for large
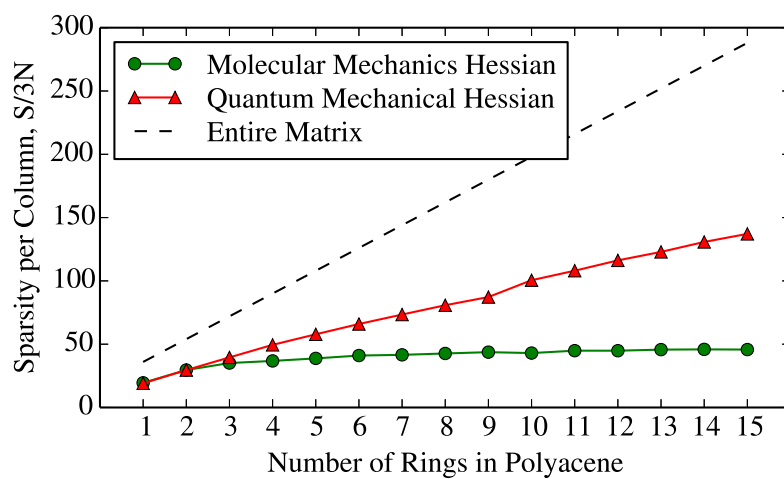
Figure 5.5: Average sparsity per column $(S/3N)$ of molecular mechanics and quantum mechanical Hessians in the basis of atomic coordinates for the series of polyacenes. (An entry in the Hessian is considered nonzero if it is greater than 10 $(\mathrm{cm}^{-1})^2$, six orders of magnitude smaller than the largest entry.) In the molecular mechanics Hessians, the average sparsity per column is roughly constant with the size of the molecule, because each atom has a roughly constant number of neighbors regardless of the size of the entire molecule.

Figure 5.6: Number of columns which must be sampled as a function of the number of rings in the polyacene to achieve a relative Frobenius norm error less than 10^{-3} in the recovered molecular mechanics Hessian. Legend entries indicate the (sparse) recovery basis, and columns are always sampled in the DCT basis with respect to the recovery basis. (Relative error is measured by averaging over ten different trials which sample different sets of random columns.)

systems. For comparison, we also recovered the Hessians in their sparsest possible basis, which is their own eigenbasis. This procedure is not practical for actual calculation since it requires knowing the entire Hessian beforehand, but it shows the best-case scenario and illustrates how the compressed sensing procedure can be improved further if an appropriate sparsifying transformation is known.

While the recovery of molecular mechanics Hessians provides a clear illustration of the scaling of our compressed sensing procedure, molecular mechanics matrix elements are not expensive to compute in comparison with rest of the linear algebra operations required to diagonalize the Hessian. Hence, from a computational standpoint, the real challenge is to apply our procedure to the computation of quantum mechanical Hessians.

As Figure 5.5 shows the sparsity S of a quantum mechanical Hessian does not necessarily scale linearly with the number of atoms N in the molecule. Figure 5.7 illustrates the cost of recovering the quantum mechanical Hessians of polyacenes using compressed sensing in a variety of sparse bases. Recovering the Hessian in the atomic coordinate basis already provides a considerable computational advantage over directly computing the entire Hessian. In fact, this curve mirrors the sparsity per column curve for quantum mechanical Hessians in Figure 5.5, consistent with our prediction that the number of sampled columns scales as $O(S/N \log N) \times OE$. More significantly, recovering the quantum mechanical Hessian in the molecular mechanics basis provides a substantial advantage over recovery in the atomic coordinates basis, reducing the number of columns which must be sampled approximately by a factor of two. This is consistent with the quantum mechanical Hessian being sparser in the approximate eigenbasis of molecular mechanics normal modes. Of course, nothing beats recovery in the exact eigenbasis, which is as sparse as possible, but requires knowing the exact Hessian in the first place.

In short, the take-home message of Figure 5.7 is that using compressed sensing to recover a quantum mechanical Hessian in its basis of molecular mechanics normal modes is a practical procedure which substantially reduces the computational cost of the procedure.
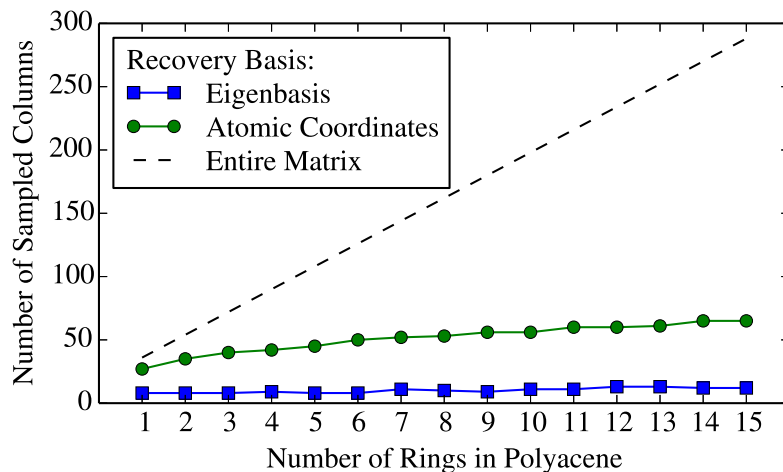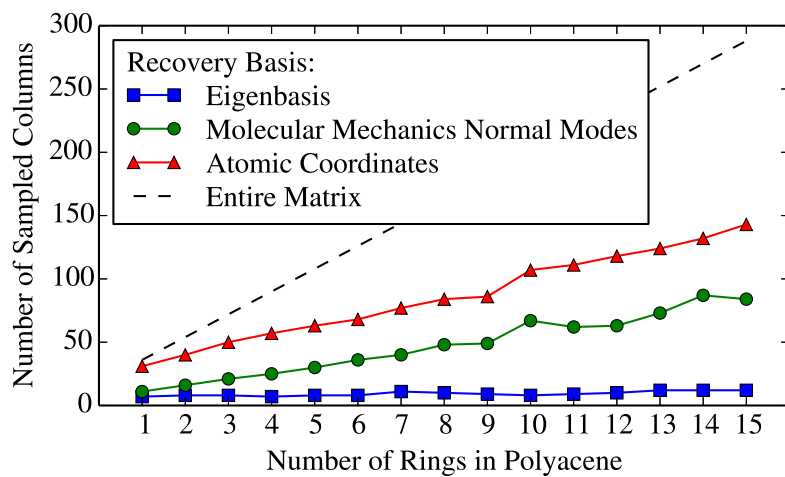
Figure 5.7: Number of columns which must be sampled as a function of the number of rings in the polyacene to achieve a relative Frobenius norm error less than 10^{-3} in the recovered quantum mechanical Hessian. Legend entries indicate the (sparse) recovery basis, and columns are always sampled in the DCT basis with respect to the recovery basis. (Relative error is measured by averaging over ten different trials which sample different sets of random columns.)

5.5 Conclusions

We have presented a new approach for calculating matrices. This method is suitable for applications where the cost of computing each matrix element is high in comparison to the cost of linear algebra operations. Our approach leverages the power of compressed sensing to avoid individually computing every matrix element, thereby achieving substantial computational savings.

When applied to molecular vibrations of organic molecules, our method results in accurate frequencies and normal modes with about 30% of the expensive quantum mechanical computations usually required, which represents a quite significant $3\times$ speed-up. Depending on the sparsity of the Hessian, our method can also improve the overall scaling of the computation. These computational savings could be further improved by using more sophisticated compressed sensing approaches, such as recovery algorithms based on belief propagation [105, 104] which offer a recovery cost directly proportional to the sparsity of the signal, and which could be easily integrated into our approach.

Our method could also be applied to other common calculations in computational chemistry, including the Fock matrix in electronic structure or the Casida matrix in linear-response time-dependent DFT [117]. Nevertheless, our method is not restricted to quantum chemistry and is applicable to many problems throughout the physical sciences and beyond. The main requirement is an *a priori* guess of a basis in which the matrix to be computed is sparse. The optimal way to achieve this requirement is problem-dependent, but as research into sparsifying transformations continues to develop, we believe our method will enable considerable computational savings in a wide array of scientific fields.

The power of compressed sensing comes from the fact that it optimizes the amount of information obtained per measurement, and hence the required number of measurements

scales with the information content being measured, or in this case calculated [172]. We believe compressed sensing can change computational chemistry for the better by focusing attention on how to avoid redundant computations and ensure that each computation delivers as much new information about a system as possible. In broad terms, the key insight of compressed sensing for computational chemistry is to *calculate only what should be calculated*.

In fact, a recent area of interest in compressed sensing is the development of dictionary learning methods that do not directly require knowledge of a sparsifying basis, but instead generate it on-the-fly based on the problem [3, 152]. We believe that combining our matrix recovery protocol with state-of-the-art dictionary learning methods may eventually result in further progress towards the calculation of scientific matrices. Beyond the problem of computing matrices, our work demonstrates that compressed sensing can be integrated into the core of computational simulations as a workhorse to reduce costs by optimizing the information obtained from each computation.

Finally, we introduced an effective method of bootstrapping low-accuracy calculations to reduce the number of high-accuracy calculations that need to be done, something which is not simple to do in quantum chemical calculations. In this new paradigm, the role of expensive high-accuracy methods is to correct the low-accuracy results, with a cost proportional to the magnitude of the required correction, rather than recalculating the results from scratch.

5.6   Computational methods

The main computational task required to implement our approach is the solution of the ℓ_1 optimization problem in equation (5.2). From the many algorithms available for this

purpose, we rely on the spectral projected gradient ℓ_1 (SPGL1) algorithm developed by van den Berg and Friedlander [184] and their freely-available implementation.

For all compressed sensing calculations in this paper, the change-of-basis matrix between the sparse basis and the measurement basis is given by the DCT matrix whose elements are given by

$$P_{ij} = \sqrt{\frac{2}{N}} \cos\left[\frac{\pi}{N}(i-1)\left(j-\frac{1}{2}\right)\right] \ , \tag{5.7}$$

with the first row multiplied by an extra factor of $1/\sqrt{2}$ to guarantee orthogonality.

For the numerical calculations we avoid explicitly constructing the Kronecker product of P with itself and instead perform all matrix multiplications in the SPGL1 algorithm directly in terms of P. This latter approach has much smaller memory requirements and numerical costs, ensuring that the compressed sensing process itself is rapid and not a bottleneck in our procedure. The condition $PAP^T = B$ is satisfied up to a relative error of 10^{-7} in the Frobenius norm (vectorial 2-norm).

In order to perform the undersampling required for our compressed sensing calculations, first the complete Hessians were calculated, then they were converted to the measurement basis, and finally they were randomly sampled by column. Quantum mechanical Hessians were obtained with the QChem 4.2 [161] software package, using density functional theory with the B3LYP exchange-correlation functional [20] and the 6-31G* basis set. Molecular mechanics Hessians were calculated using the the MM3 force field [6] and the open-source package Tinker 6.2.

5.7 Appendix A: Further Comparison of Vibrational Frequencies

One reasonable question is whether the compressed sensing procedure outlined in the main article accurately recovers changes in the vibrational modes (i.e. the eigenvectors of the Hessian matrix) or whether it is helpful only in recovering changes in the vibrational frequencies (i.e. only the diagonal elements on the Hessian matrix). To address this question, we evaluated the true quantum mechanical (QM) vibrational frequencies along the vibrational modes obtained by the cheap molecular mechanics (MM) calculation and compared the results with compressed sensing. The results for anthracene are shown in Figure 5.8. In particular, the figure compares the MM frequencies, the QM frequencies along the MM modes, and the compressed sensing frequencies with 35% sampling of the QM calculations according to the procedure described in the main article. (Note that the two plots present the same data with different frequency scales on the y-axis.)

The top plot shows that computing the QM frequencies along the MM modes provides a large improvement on the MM frequencies themselves, as expected since the eigenvalue calculation is quantum mechanical in nature and only the eigenvectors are approximate. However, substantial errors up to 100 cm^{-1} remain in some of the frequencies so this approach does not suffice for chemical accuracy. In contrast, as the bottom plot shows, the compressed sensing frequencies are nearly exact, and are far more accurate than the QM frequencies computed along the MM modes. This result provides concrete numerical evidence that compressed sensing accurately recovers changes not only in the eigenvalues (i.e. diagonal elements of the Hessian matrix) but also in the vibrational modes themselves (i.e. the eigenvectors). As can be seen, only the compressed sensing approach, with accurate
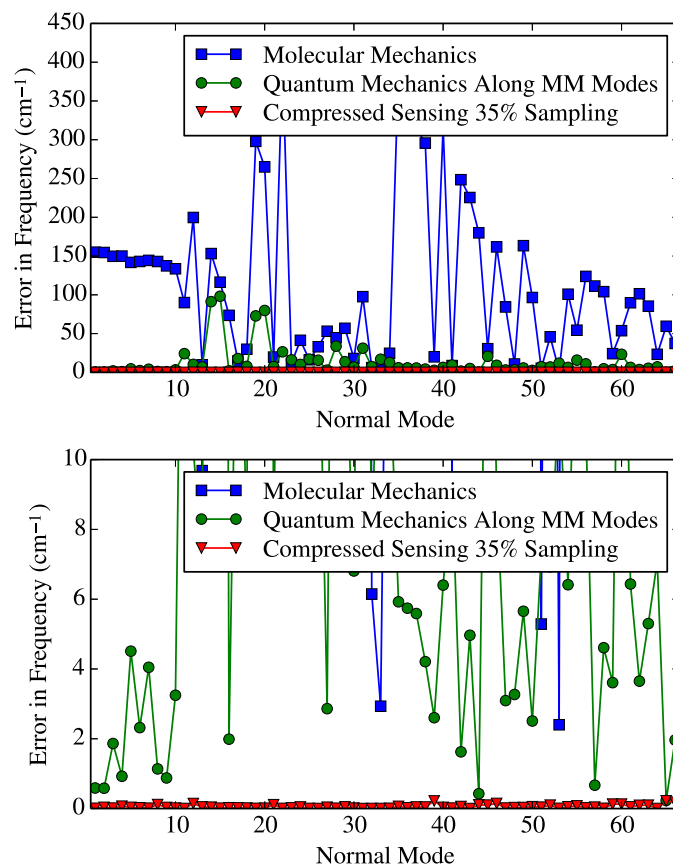
Figure 5.8: The figure compares the MM frequencies, the QM frequencies along the MM modes, and the compressed sensing frequencies with 35% sampling of the QM calculations according to the procedure described in the main article for anthracene. (Note that the two plots present the same data with different frequency scales on the y-axis.) The QM frequencies along the MM modes are much more accurate than the MM frequencies themselves, but only the compressed sensing calculation provides acceptable chemical accuracy.

recovery of both vibrational modes and frequencies, suffices for chemical accuracy.

We also point out that as a method, computing the QM frequencies along the MM modes is based on an uncontrolled approximation, and the method can fail if the normal modes are predicted badly. With compressed sensing, a bad prediction of the normal modes is simply reflected in a more expensive calculation, but the results would still be accurate to the QM level.

5.8   Appendix B: Convergence Criterion

For practical implementation of the compressed sensing procedure (steps 1–7 in the main article), an important requirement is a convergence criterion; namely, how do you know when the procedure is done? In particular, without comparing to a full reference calculation, how do you know when enough columns of the Hessian have been sampled to achieve accurate vibrational modes and frequencies?

To address this question, we repeatedly applied the compressed sensing procedure (steps 1–7 in the main article) to the quantum mechanical Hessian matrix of anthracene, and each time we sampled one additional column. For each additional column sampled, we computed the relative Frobenius norm (vectorial 2-norm) error against the full reference calculation,

$$\text{Overall Error}_n = \frac{||H_n - H_{\text{ref}}||_2}{||H_{\text{ref}}||_2}, \tag{5.8}$$

where H_n is the Hessian recovered via compressed sensing with n columns sampled in the measurement basis. We also computed the relative Frobenius norm error against the calculation in the previous step with one less sampled column,

$$\text{Successive Error}_n = \frac{||H_n - H_{n-1}||_2}{||H_n||_2}. \tag{5.9}$$
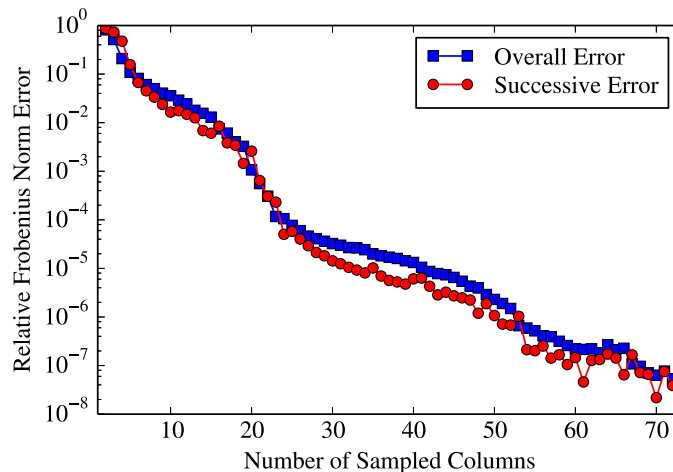
Figure 5.9: The figure shows both the overall error and the successive error in the Hessian of anthracene obtained via the compressed sensing procedure described in the main article as additional columns are sampled. Because the two errors track each other closely, the successive error can be used as a basis for developing a convergence criterion in the absence of a full reference calculation.

Figure 5.9 shows both types of error plotted as a function of the number of sampled columns n. As the figure shows, the two errors track each other closely (never differing by more than an order of magnitude), but the successive error can be computed without access to a full reference calculation. Small successive errors indicate that the Hessian matrix is no longer changing when additional columns are sampled, and show that the calculation has converged. Hence, just as in many other numerical methods, the successive error provides a measure for developing a convergence criterion. For example, to assure robustness, one might demand small successive errors over three or four consecutive steps before declaring the calculation complete.

Acknowledgments

Chapter 6

Compressed Sensing Reduces CPU Time Needed to Compute Molecular Hessians

Abstract

This paper describes the implementation of compressed sensing into Q-Chem, a commercial electronic structure software package, to more efficiently compute Hessian matrices, whose diagonalization yields the vibrational normal modes and frequencies of a molecule. The key advantage of compressed sensing is that only a subset of the columns of the Hessian matrix needs to be computed, resulting in a large savings in CPU hours. However, the columns must be computed not in the usual basis where each atomic nucleus is independently displaced in three Cartesian directions, but rather in a basis consisting of collective displacements of many nuclei at once. This paper outlines the technique and applies it to the vibrational frequencies of a linear polyacene series computed with density functional theory using the M06-2X functional in the double-zeta 6-31G(d,p) basis. For this series of molecules, compressed sensing reduces by 50% the number of columns which must be computed and correspondingly decreases the required number of CPU hours by over 40%. Future improvements to the procedure will likely reduce the number of CPU hours even

further.

6.1 Introduction

Many scientific applications involve many calculations, followed by compressing or consolidating the data to obtain new information. An emerging paradigm is instead to "measure what should be measured," which is the idea that the number of calculations should directly relate to the information content present in the data [172]. One way to implement this paradigm is through a state-of-the-art signal processing method known as compressed sensing: first, a sparse representation of the data is found and second, this data is recovered through a number of calculations that is roughly proportional to its sparsity [34, 57, 35].

In quantum chemistry, one of the most fertile applications for this new paradigm is the computation of Hessian matrices, as this is one of the most expensive computations routinely performed by chemists to determine whether a particular geometry corresponds to the ground state of a molecule or the transition state of a reaction [190]. The Hessian matrix of a molecule contains the mass-weighted second derivatives of its electronic energy with respect to nuclear displacements, and diagonalization of this matrix provides its vibrational normal modes (as the eigenvectors) and frequencies (as the square root of the eigenvalues) [52].

In a previous paper [154], we developed and simulated the use of compressed sensing to reduce the cost of computing Hessian matrices. This paper follows up with an implementation in the Q-Chem electronic software package [161], translating the simulations of the previous paper into real reductions in computational cost as measured by fewer CPU hours.

6.2 Computational Methods

In quantum chemistry methods, derivatives of the electronic energy may be computed analytically [197, 66, 141, 65, 74, 54], by differentiating the energy expression, or numerically, by using finite differences. This paper focuses on so-called mixed derivatives, where gradients of the electronic energy are computed analytically but second derivatives are computed via finite differences of the gradients. Mixed derivatives are popular because analytic second derivatives are typically much more challenging to compute than analytic gradients. As a consequence of the Hellmann-Feynman theorem [77, 62], analytic gradients may be obtained without computing the first derivatives of the molecular orbitals with respect to nuclear displacements, but this fortunate state of affairs no longer holds for second derivatives, where the molecular orbital derivatives must be obtained by solving memory-intensive coupled-perturbed equations. Because of these memory limitations, many popular electronic structure methods use mixed derivatives, including Møller-Plesset second-order perturbation theory (MP2) [126], coupled-cluster singles-doubles (CCSD) [143], and many density functionals including the popular Minnesota functional M06-2X [202]. Hence we expect our compressed sensing implementation to enjoy broad applicability.

Consider a configuration of N nuclei expressed by the mass-weighted coordinates $\boldsymbol{q} = \{q_1, \ldots, q_{3N}\}$, and suppose analytic first derivatives of the energy, $\dfrac{\partial E}{\partial q_i}$, are available for each q_i. Then, using mixed derivatives, elements of the Hessian matrix are computed as

$$\frac{\partial^2 E(\boldsymbol{q})}{\partial q_i \partial q_j} = \frac{1}{2h} \left\{ \frac{\partial E}{\partial q_i} \left(\boldsymbol{q} + h\widehat{q_j}\right) - \frac{\partial E}{\partial q_i} \left(\boldsymbol{q} - h\widehat{q_j}\right) \right\} , \tag{6.1}$$

where $\widehat{q_j}$ is a unit vector along the jth coordinate and h is the step size used in the finite difference calculation. It is worth taking a moment to unpack equation (6.1). A single gradient calculation at a particular point $(\cdot)$ yields $\dfrac{\partial E}{\partial q_i}(\cdot)$ for all q_i. Hence, two gradient

calculations, $\frac{\partial E}{\partial q_i}(\boldsymbol{q} + h\widehat{q_j})$ and $\frac{\partial E}{\partial q_i}(\boldsymbol{q} - h\widehat{q_j})$, are required to obtain the jth column of the Hessian matrix. Since the Hessian matrix contains $3N$ columns, it follows that a total of $6N$ gradient calculations are required to recover the entire matrix.

As discussed in our previous paper [154], the key advantage afforded by compressed sensing is that it enables the full recovery of a Hessian matrix by computing only a subset of its columns and hence only a subset of the gradient calculations. Since the total CPU time is directly proportional to the number of gradients which must be computed, the application of compressed sensing yields a direct reduction in CPU time. Because compressed sensing is a technique for recovering sparse vectors and matrices from undersampled data, the first requirement is locating a basis in which the Hessian matrix is sparse. One choice is a basis of approximate normal modes, as may be obtained from a cheap Hartree-Fock calculation in a minimal STO-3G basis, in which the more exact Hessian is expected to be nearly diagonal with a few small off-diagonal entries. These approximate normal modes are transformed according to the discrete cosine transform (DCT) and only a subset of the columns of the Hessian needs to be computed in this DCT basis. From this subset of columns, compressed sensing is able to recover the complete sparse Hessian in the approximate normal mode basis. Once the complete Hessian is recovered, it may be transformed back into the atomic coordinate basis and diagonalized to obtain the vibrational modes and frequencies.

Hence, the key requirement for compressed sensing is that second derivatives must be computed not in the usual basis where each atomic nucleus is independently displaced in three Cartesian directions, but rather in a basis consisting of collective displacements of many nuclei at once–in particular, the discrete cosine transform basis of the approximate normal modes. The advantage is that, in this DCT basis, only a subset of the columns need to be computed [154], which translates directly into a subset of the gradients and hence a

net reduction of CPU time.

These changes result in the following procedure for computing Hessian matrices, in which we use upper-case letters Q_i to denote coordinates in the DCT basis, while we continue to use lower-case letters q_i to denote mass-weighted coordinates in the atomic nuclei basis. This procedure, which can be toggled on and off by setting the value of the new Q-Chem keyword CS_HESS to 1 and 0 respectively, is as follows:

1. For a *subset* of directions $\widehat{Q_j}$ in the DCT basis, compute the analytic gradients $\frac{\partial E}{\partial Q_i}\left(\boldsymbol{Q} + h\widehat{Q_j}\right)$ and $\frac{\partial E}{\partial Q_i}\left(\boldsymbol{Q} - h\widehat{Q_j}\right)$. The number of directions $\widehat{Q_j}$ chosen is equal to three times the value of the user-specified keyword CS_ATOMS_PER_ITER.

2. Assemble this subset of analytic gradients to yield columns of the Hessian matrix,

$$\frac{\partial^2 E(\boldsymbol{Q})}{\partial Q_i \partial Q_j} = \frac{1}{2h}\left\{\frac{\partial E}{\partial Q_i}\left(\boldsymbol{Q} + h\widehat{Q_j}\right) - \frac{\partial E}{\partial Q_i}\left(\boldsymbol{Q} - h\widehat{Q_j}\right)\right\} . \qquad (6.2)$$

Note that we have *all* the Q_i derivatives (i.e. all the rows) but only a subset of the Q_j derivatives (i.e. only a subset of the columns).

3. Apply compressed sensing, as fully described in our earlier paper [154], to attempt to recover the full Hessian matrix in the sparse basis of approximate normal modes. To check for convergence, repeat the compressed sensing procedure using one less column of the matrix in step 2. The results are converged when the relative Frobenius norm error between the two recovered matrices is less than $10^{-\text{CS\_TOL}}$, where CS_TOL is a user-specified keyword. In Q-Chem, the compressed sensing procedure is done by the new function bpdn which solves the basis-pursuit denoising problem using the algorithm described in ref. [184].

4. If the results are not converged, return to Step 1 and choose another subset of directions $\widehat{Q_j}$ along which to compute analytic gradients, thus yielding more columns of

the Hessian matrix. If the results are converged, convert the recovered matrix back into the atomic nuclei basis $\dfrac{\partial^2 E(\boldsymbol{q})}{\partial q_i \partial q_j}$ and diagonalize it to obtain the vibrational normal modes (as the eigenvectors) and vibrational frequencies (as the square root of the eigenvalues).

6.3   Results and Discussion

This section compares the compressed sensing approach described above to the traditional approach for computing Hessian matrices. In particular, we examine how many columns of the Hessian must be computed to recover the vibrational frequencies of a molecule accurately via compressed sensing and we quantify how much CPU time is saved compared to the traditional approach as a result.

As a model system, we compute the vibrational frequencies of the linear polyacene series ranging from benzene to pentacene via density functional theory with the Minnesota M06-2X functional [202] in a double-zeta 6-31G(d,p) basis. The M06-2X functional is popular in organic chemistry applications, and elements of the Hessian matrix are generally computed via the mixed approach described above, using analytical gradients and numerical second derivatives, as fully analytical second derivatives are not available. Hence, the M06-2X functional is an ideal test case for the compressed sensing method outlined above. As a rapid method for obtaining approximate vibrational modes, we use Hartree-Fock theory in a minimal STO-3G basis, as the M06-2X/6-31G(d,p) Hessian is expected to be reasonably sparse in the basis of these approximate modes. In all calculations using the compressed sensing approach, the keyword CS_TOL is set to 4, resulting in a tolerance of 10^{-4} for determining when the compressed sensing process has converged.

All vibrational frequency calculations were performed on the Edison computational
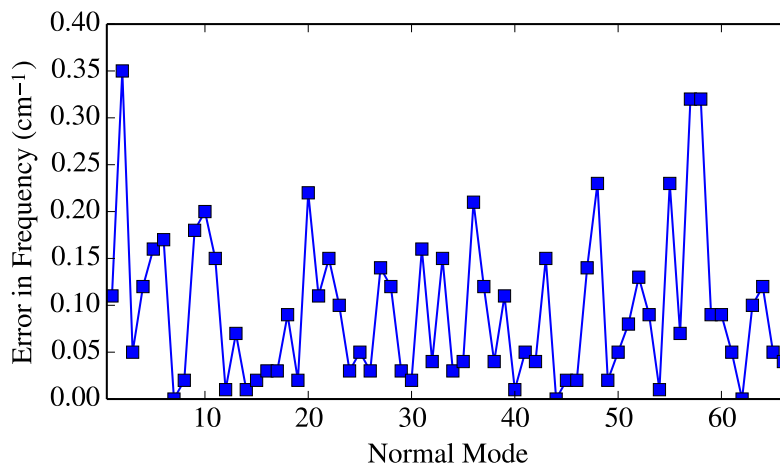
Figure 6.1: This figure shows the error in the vibrational frequencies of anthracene obtained via compressed sensing. All of the frequencies are recovered to within 0.35 cm^{-1}, which is well within the limit of chemical accuracy.

system of the National Energy Research Scientific Computing Center. In each calculation, the total number of CPU hours was quantified on a computational node containing two 12-core Intel "Ivy Bridge" processors at 2.4 GHz and 64 GB DDR3 1866 MHz of memory.

We take up two questions. First, how accurate are the frequencies obtained via compressed sensing as compared to the traditional approach? Second, how much CPU time is saved by the compressed sensing approach?

To answer the first question, Figure 6.1 illustrates the error in the vibrational frequencies of anthracene obtained via compressed sensing (steps 1–4 in the previous section) as compared to the traditional approach. As the plot shows, all of the vibrational frequencies are recovered to within 0.35 cm^{-1}, well within a few wavenumbers which constitutes the generally accepted benchmark of chemical accuracy. Similarly accurate results were obtained for benzene and naphthalene.

Figure 6.2 illustrates the error in vibrational frequencies for pentacene obtained via

Figure 6.2: This figure shows the absolute error in the vibrational frequencies of pentacene obtained via compressed sensing. Most of the frequencies are recovered to within 1 cm^{-1}, but the highest ones corresponding to C–H stretching exhibit larger errors around 5 cm^{-1}.
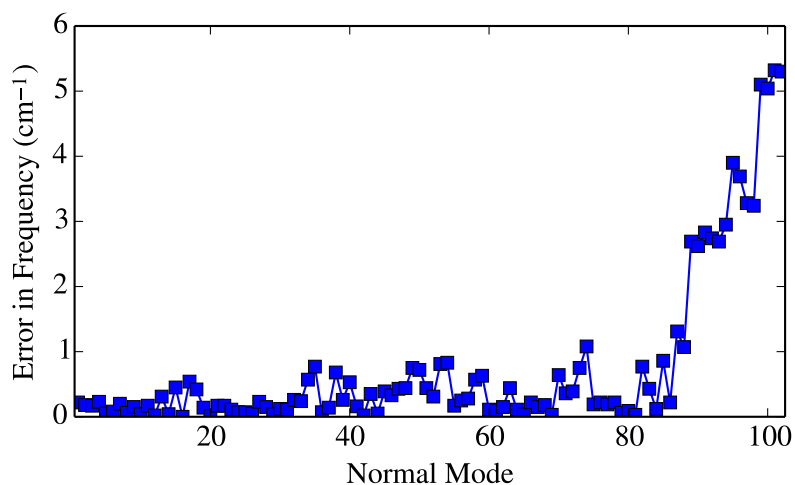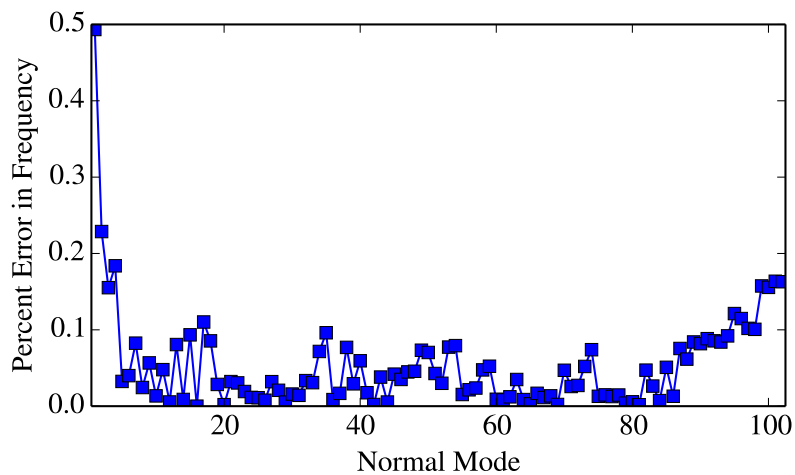

Figure 6.3: This figure shows the relative error in the vibrational frequencies of pentacene obtained via compressed sensing, expressed as a percent. All of the frequencies are recovered to within 0.5%, including the highest ones corresponding to C–H stretching.

compressed sensing. Most of the frequencies are recovered accurately to within 1 cm^{-1}, but the highest frequencies (corresponding to the stretching of C–H bonds) exhibit larger errors up to just above 5 cm^{-1}. A similar result was also obtained for tetracene, whereby some of the highest frequencies corresponding to C–H stretches exhibited errors around 5 cm^{-1}. Figure 6.3 replots this error as the *relative* error in frequency, expressed as percent error, and shows that all frequencies are recovered accurately to within 0.5%, and that the relative error does not display any notable increase at the highest frequencies. Nevertheless, there are two possible sources of this error. First, the compressed sensing process may not have fully converged to resolve vibrational frequencies to within 1 cm^{-1}, and hence sampling more columns of the Hessian in the DCT basis is necessary. Second, the derivatives have a small amount of variability when they are computed in a different basis (in this case in a DCT basis rather than the standard atomic basis), owing to the discretization process whereby the derivatives are approximated as differences between grid points.

By computing the full Hessian in the DCT basis, we were able to separate these two sources of error, and the results for pentacene are shown in Figure 6.4. The plot clearly shows that nearly all of the error comes not from the compressed sensing process but from the fact that the Hessian is being sampled in the DCT basis rather than the atomic basis. From a different perspective, this "error" is not really error after all, since it is natural variability owing to the discretization process whereby the derivatives are approximated via finite differences. Such variability could be minimized by using a larger stencil, for example.

As for the second question, regarding how much computational effort is saved by compressed sensing, Figure 6.5 compares the number of columns which must be computed in the compressed sensing method to the total number of columns which must be computed in the traditional approach (which is equal to $3N$, where N is the number of atoms). Across the
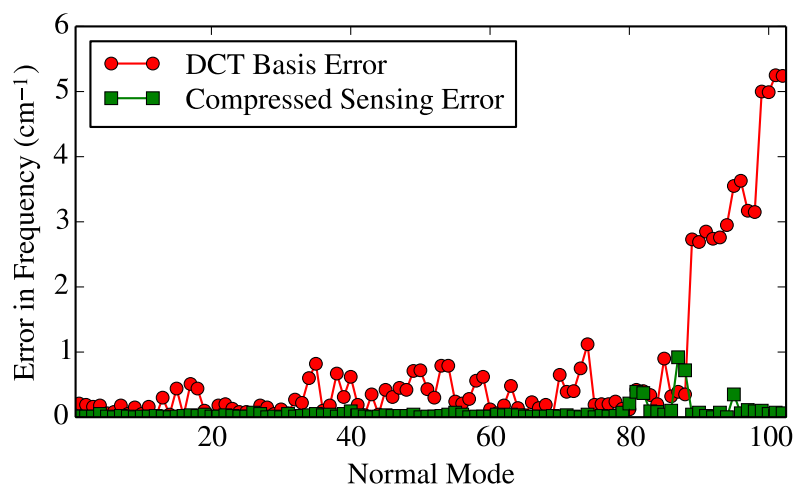
Figure 6.4: This figure decomposes the error in the vibrational frequencies of pentacene into two parts: DCT basis error (which is natural variability in the finite difference approximation when derivatives are computed in a different basis) and compressed sensing error (which arises when not enough columns of the Hessian have been sampled). The plot shows that compressed sensing has converged to within chemical accuracy, and nearly all of the "error" is simply the result of taking derivatives in a different basis.
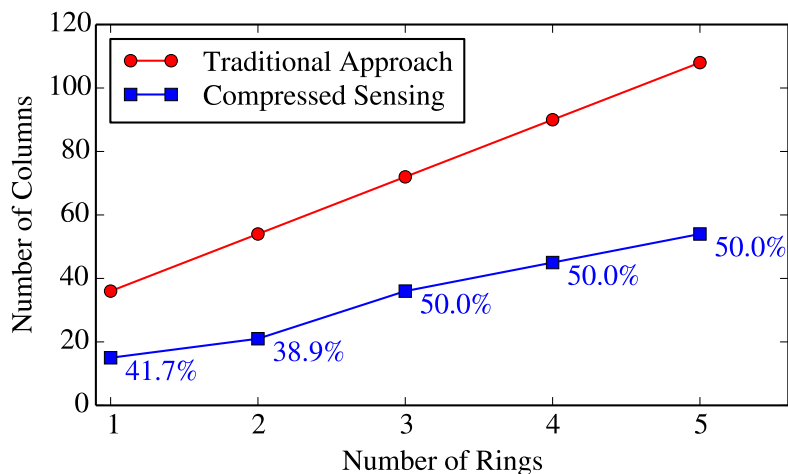
Figure 6.5: This figure shows how many columns of the Hessian must be sampled in both the traditional approach and compressed sensing for the linear polyacene series. Compressed sensing yields a savings of approximately 50% in the number of columns that must be computed.

polyacene series from benzene to pentacene, the number of columns of the Hessian which must be sampled for accurate recovery of vibrational frequencies consistently hovers at or slightly below 50%, which corresponds to cutting in half the number of analytic gradient calculations which must be done.

Figure 6.6 shows that the reduction in the number of analytic gradient calculations which must be done translates nicely into a reduction of CPU hours required. The number of CPU hours required for compressed sensing hovers at or below 60% of the number of CPU hours required in the traditional approach. The reason approximately 60% of the CPU hours are required when only 50% of the analytic gradients are being computed stems from the computational cost of the compressed sensing process itself (i.e. the BPDN procedure). This cost can be almost completely eliminated by increasing the number of columns which are computed at a time before the compressed sensing process is performed. For example,
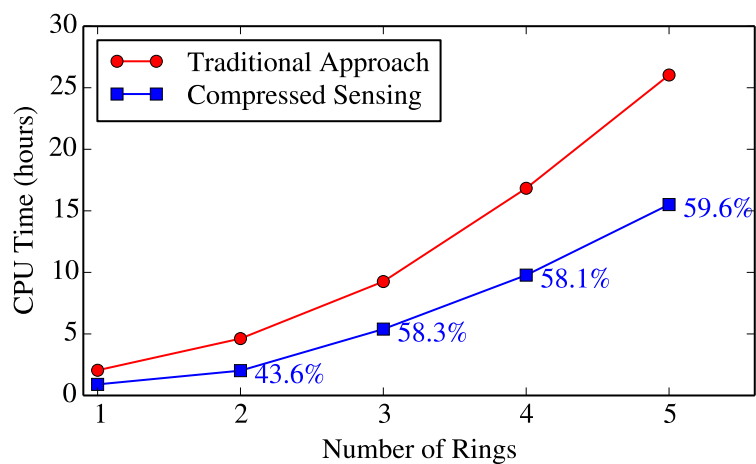
Figure 6.6: This figure quantifies the number of CPU hours required to compute the vibrational modes and frequencies of linear polyacenes in both the traditional approach and compressed sensing. The savings in CPU hours for compressed sensing closely mirror the reduction in the number of columns of the Hessian that must be computed, with a small additional cost of CPU hours to perform the compressed sensing process itself.

if it is known that approximately 50% of the columns are required for accurate recovery of the vibrational frequencies, then there is no need to perform compressed sensing and check for convergence before nearly 50% of the columns have been computed. By choosing more strategically when compressed sensing is performed, the percentage of CPU hours required can be reduced asymptotically towards the percentage of columns which must be computed.

To enable this possibility, we are introducing a new feature into Q-Chem called FastCS. By setting the keyword CS_FAST to a number between 0 and 100, it is possible to specify the percentage of columns of the Hessian matrix which are computed *a priori* before the compressed sensing process (i.e. the BPDN procedure) is performed for the first time. In the examples above, for example, it would be prudent to set CS_FAST to 40 or 50.

In conclusion, compressed sensing provides a substantial reduction in the number of CPU hours required to compute the vibrational normal modes and frequencies of molecules. Further computations will investigate the performance of compressed sensing with other electronic structure methods such as MP2 and CCSD, as these methods also employ the mixed approach of analytic gradients and numerical second derivatives. In addition, computations will progress to larger molecules to investigate how the reduction in CPU hours scales with the number of atoms in the molecule. Finally, by varying the step size used in the finite difference calculations, we will investigate and attempt to reduce the variability in the highest vibrational frequencies with respect to the basis in which the Hessian is computed. Taken together, these further computations should increase the accuracy and breadth of applicability of compressed sensing, as well as provide further reductions in CPU hours to obtain vibrational normal modes and frequencies.

Acknowledgements

Chapter 7

Q-Chem Implementation of Compressed Sensing for the Computation of Analytical Second Derivatives

Abstract

Q-Chem is a commercial quantum chemistry package used by many theoretical and experimental chemists to compute molecular energies, geometries, vibrational and electronic spectra, and other properties. This chapter describes the implementation of compressed sensing into Q-Chem for the more efficient computation of Hessian matrices, whose diagonalization yields the vibrational normal modes and frequencies of a molecule. Unlike the previous chapter, which focuses on so-called mixed derivatives (analytic gradients and numerical second derivatives), this chapter describes the implementation of compressed sensing with fully analytical second derivatives, as these derivatives are often more efficient to compute when they are available for a particular electronic structure method. While the method has been fully implemented, unlike with mixed derivatives, we have not yet observed a reduction in the number of CPU hours required to compute analytical second

derivatives with compressed sensing. Understanding why this is the case is the subject of ongoing investigation, but we remain optimistic that the approach could prove useful in certain quantum chemistry packages depending on how they implement analytical second derivatives.

7.1 Introduction

The chapter begins with a mathematical presentation of first and second derivative methods in quantum chemistry, with an emphasis on the Hartree-Fock theory. The key pedagogical reference for wavefunction methods is a monograph by Yamaguchi *et al.* [197], with appropriate modifications made to the Hartree-Fock formalism [66, 141, 65] to account for density functional theory available in references [74, 54]. This first half of the chapter ends with the step-by-step procedure used by Q-Chem [161] to compute the second derivatives of the molecular energy needed to obtain the vibrational modes and frequencies, detailing which functions are responsible for computing the various components. Further references [99, 68, 129] discuss certain aspects of the implementation of derivative methods in Q-Chem [161]. We also limit the discussion to restricted closed-shell computations; the generalization to unrestricted computations is straightforward but more cumbersome.

The second half of the chapter explains how this Hessian computation procedure must be modified to take advantage of the compressed sensing methods described in the previous chapter and in ref. [154]. The most expensive parts of the Hessian matrix are the components which depend on the first derivative of the density matrix, which require solving a new set of coupled-perturbed equations for each of the $3N$ derivatives that are required for a molecule containing N atoms. The key advantage afforded by compressed sensing is that it enables dramatic undersampling below $3N$ in the number of coupled-perturbed equations

which must be solved, resulting in a reduction in the amount of computational time and parallelization needed to compute Hessian matrices. However, to take advantage of this undersampling, the coupled-perturbed equations must be solved not in the usual basis where each atomic nucleus is independently displaced in three Cartesian directions, but rather in an arbitrary basis consisting of collective displacements of many nuclei at once. This yields an undersampling of the columns of the Hessian in a basis of collective displacements, from which the full Hessian can be recovered in a sparse basis according to the compressed sensing methods described in the previous chapter and in ref. [154]. The second half of this chapter includes a step-by-step procedure for how compressed sensing is implemented into Q-Chem, with a detailed accounting of which new functions are responsible for the various modifications to the standard procedure.

7.2   Integrals in Quantum Chemistry

Consider a set of atomic basis functions $\{\chi_\mu\}$ from which we construct molecular orbitals $\{\phi_i\}$ according to

$$\phi_i = \sum_\mu^{\text{AO}} C_\mu^i \chi_\mu, \tag{7.1}$$

where the C_μ^i are the molecular orbital expansion coefficients and AO serves as a reminder that the sum is over atomic basis functions. For simplicity, we assume all atomic basis functions and molecular orbitals are real; the generalization to complex functions is straightforward. In the atomic orbital basis, we define the overlap integral

$$S_{\mu\nu} = \int \chi_\mu(1)\chi_\nu(1)d\tau_1, \tag{7.2}$$

and the single-electron integral

$$h_{\mu\nu} = \int \chi_\mu(1) \left(-\frac{1}{2}\nabla_1^2 - \sum_A \frac{Z_A}{r_{1A}} \right) \chi_\nu(1) d\tau_1, \tag{7.3}$$

where Z_A is the charge on nucleus A and r_{1A} is the distance from electron 1 to nucleus A, as well as the two-electron integral

$$(\mu\nu|\rho\sigma) = \int\int \chi_\mu(1)\chi_\nu(1)\frac{1}{r_{12}}\chi_\rho(2)\chi_\sigma(2)d\tau_1 d\tau_2, \tag{7.4}$$

where r_{12} is the distance from electron 1 to electron 2.

We make similar definitions in the molecular orbital basis, defining the overlap integral

$$S_{ij} = \int \phi_i(1)\phi_j(1)d\tau_1 = \sum_{\mu\nu}^{AO} C_\mu^i C_\nu^j S_{\mu\nu}, \tag{7.5}$$

the single-electron integral

$$h_{ij} = \int \phi_i(1) \left(-\frac{1}{2}\nabla_1^2 - \sum_A \frac{Z_A}{r_{1A}} \right) \phi_j(1)d\tau_1 = \sum_{\mu\nu}^{AO} C_\mu^i C_\nu^j h_{\mu\nu}, \tag{7.6}$$

and the two-electron integral

$$(ij|kl) = \int\int \phi_i(1)\phi_j(1)\frac{1}{r_{12}}\phi_k(2)\phi_l(2)d\tau_1 d\tau_2 = \sum_{\mu\nu\rho\sigma}^{AO} C_\mu^i C_\nu^j C_\rho^k C_\sigma^l (\mu\nu|\rho\sigma), \tag{7.7}$$

where we have used the relationship between atomic and molecular orbitals in equation (7.1) to relate the integrals in the MO basis to the integrals in the AO basis.

In terms of these integrals, we define the Fock matrix in the MO basis as

$$F_{ij} = h_{ij} + 2\sum_k^{DO} \left\{ (ij|kk) - \frac{1}{2}(ik|ik) \right\} \tag{7.8}$$

where DO signifies that the sum is over all doubly-occupied molecular orbitals. Using the relationships between the integrals in the MO basis and the integrals in the AO basis in equations (7.6) and (7.7), the Fock matrix may also be expressed in the AO basis as

$$F_{\mu\nu} = h_{\mu\nu} + \sum_{\rho\sigma}^{AO} P_{\rho\sigma} \left\{ (\mu\nu|\rho\sigma) - \frac{1}{2}(\mu\rho|\nu\sigma) \right\}, \tag{7.9}$$

where $P_{\mu\nu}$ is the density matrix (in the atomic orbital basis) defined by

$$P_{\mu\nu} = 2 \sum_{i}^{\text{DO}} C_{\mu}^{i} C_{\nu}^{i}. \tag{7.10}$$

7.3 Electronic Energy

In a restricted Hartree-Fock calculation, in which all molecular orbitals are doubly-occupied, the total electronic energy E is given by

$$E = 2 \sum_{i}^{\text{DO}} h_{ii} + 2 \sum_{ij}^{\text{DO}} \left\{ (ii|jj) - \frac{1}{2}(ij|ij) \right\}. \tag{7.11}$$

Using the relationships between the MO basis and the AO basis in equations (7.6) and (7.7), as well as the definition of the density matrix in equation (7.10), the total energy may be expressed in the AO basis as [197]

$$E = \sum_{\mu\nu}^{\text{AO}} P_{\mu\nu} h_{\mu\nu} + \frac{1}{2} \sum_{\mu\nu\rho\sigma}^{\text{AO}} P_{\mu\nu} P_{\rho\sigma} \left\{ (\mu\nu|\rho\sigma) - \frac{1}{2}(\mu\rho|\nu\sigma) \right\}. \tag{7.12}$$

Our goal in the first part of this chapter is to explain how derivatives of the electronic energy, as expressed in equations (7.11) and (7.12), are computed.

7.4 First Derivative of the Molecular Orbitals

We have the choice of differentiating the electronic energy in the MO basis, beginning from equation (7.11), or in the AO basis, beginning from equation (7.12). For the first derivative, it is simpler to work in the MO basis. To understand the logic involved in taking derivatives of equation (7.11), consider a molecule in its ground state and imagine perturbing its nuclei. The values of the one-electron integrals, h_{ii}, and two-electron integrals, $(ii|jj)$ and $(ij|ij)$, change for two reasons. Most obviously, they change because they contain explicit dependence on the nuclear coordinates. Slightly more subtle is the fact that the

molecular orbitals i and j themselves also change when the nuclei are perturbed. Thus, a derivative computation involves two aspects: derivatives of the one-electron and two-electron integrals, with the molecular orbitals held fixed, and derivatives of the molecular orbital coefficients C_μ^i themselves.

Consider the perturbation of molecular orbital i by the displacement of a nucleus by Δx. We can express this as a perturbation of the expansion coefficients C_μ^i,

$$C_\mu^{i,\text{pert}} = C_\mu^i + \Delta x \frac{\partial C_\mu^i}{\partial x} \tag{7.13}$$

We then make the *ansatz* that the changes in the molecular orbital coefficients C_μ^i may be expressed as linear combinations of the *unperturbed* molecular orbitals,

$$C_\mu^{i,\text{pert}} = C_\mu^i + \Delta x \sum_m^{\text{MO}} U_{mi}^x C_\mu^m, \tag{7.14}$$

where the U_{mi}^x are the expansion coefficients and MO denotes that the sums are over *all* molecular orbitals, not just those which are doubly-occupied.

Comparing equations (7.13) and (7.14) reveals that the expansion coefficients U_{mi}^x are defined according to [197]

$$\frac{\partial C_\mu^i}{\partial x} = \sum_m^{\text{MO}} U_{mi}^x C_\mu^m. \tag{7.15}$$

Equation (7.15) expresses the effects of perturbing the molecular orbitals in terms of the unperturbed molecular orbitals.

7.5   First Derivative of the Electronic Energy

In order to differentiate the electronic energy as expressed in the MO basis in equation (7.11), we require the derivatives of the overlap S_{ij}, one-electron h_{ij}, and two-electron $(ij|kl)$ integrals. To do this, we relate these integrals in the MO basis to integrals in the AO basis

118

and apply the product rule. For example, the orbital overlap S_{ij} may be expressed as the sum

$$S_{ij} = \sum_{\mu\nu}^{\text{AO}} C_\mu^i C_\nu^j S_{\mu\nu}, \tag{7.16}$$

and differentiating this expression yields

$$\frac{\partial S_{ij}}{\partial x} = \sum_{\mu\nu}^{\text{AO}} C_\mu^i C_\nu^j \frac{\partial S_{\mu\nu}}{\partial x} + \sum_{\mu\nu}^{\text{AO}} \left(\frac{\partial C_\mu^i}{\partial x} C_\nu^j + C_\mu^i \frac{\partial C_\nu^j}{\partial x} \right) S_{\mu\nu}, \tag{7.17}$$

where the first term

$$S_{ij}^x = \sum_{\mu\nu}^{\text{AO}} C_\mu^i C_\nu^j \frac{\partial S_{\mu\nu}}{\partial x} \tag{7.18}$$

is known as a *core* overlap integral derivative, which is purely due to changes in the atomic orbitals $\{\chi_\mu\}$ themselves, while the second term accounts for changes in the molecular orbitals. Using equation (7.15) for the derivatives of the molecular orbital coefficients, as well as equation (7.16), the derivative of the overlap integral in the MO basis may be written as

$$\frac{\partial S_{ij}}{\partial x} = S_{ij}^x + \sum_m^{\text{MO}} \left(U_{mi}^x S_{mj} + U_{mj}^x S_{im} \right). \tag{7.19}$$

By exactly the same procedure, the first derivatives of the one-electron integrals may be computed as

$$\frac{\partial h_{ij}}{\partial x} = h_{ij}^x + \sum_m^{\text{MO}} \left(U_{mi}^x h_{mj} + U_{mj}^x h_{im} \right), \tag{7.20}$$

where

$$h_{ij}^x = \sum_{\mu\nu}^{\text{AO}} C_\mu^i C_\nu^j \frac{\partial h_{\mu\nu}}{\partial x} \tag{7.21}$$

is the *core* one-electron integral derivative, which accounts for changes in the atomic orbitals and the molecular Hamiltonian, while the second term containing U_{mi}^x and U_{mj}^x account for changes in the molecular orbitals.

119

By an analogous process, the first derivatives of the two-electron integrals are given by

$$\frac{\partial(ij|kl)}{\partial x} = (ij|kl)^x + \sum_m^{\text{MO}} \left\{ U_{mi}^x(mj|kl) + U_{mj}^x(im|kl) + U_{mk}^x(ij|ml) + U_{ml}^x(ij|km) \right\}, \quad (7.22)$$

where

$$(ij|kl)^x = \sum_{\mu\nu\rho\sigma}^{\text{AO}} C_\mu^i C_\nu^j C_\rho^k C_\sigma^l \frac{\partial(\mu\nu|\rho\sigma)}{\partial x} \quad (7.23)$$

is the *core* two-electron integral derivative, which accounts for changes in the atomic orbitals and the molecular Hamiltonian, while the second term accounts for changes in the molecular orbitals.

We are finally ready to differentiating the electronic energy in equation (7.11), obtaining

$$\frac{\partial E}{\partial x} = 2\sum_i^{\text{DO}} \frac{\partial h_{ii}}{\partial x} + 2\sum_{ij}^{\text{DO}} \left\{ \frac{\partial(ii|jj)}{\partial x} - \frac{1}{2}\frac{\partial(ij|ij)}{\partial x} \right\}. \quad (7.24)$$

Substituting the derivatives of the one-electron and two-electron integrals in equations (7.20) and (7.22), and making use of the definition of the Fock matrix in equation (7.8), ultimately yields

$$\frac{\partial E}{\partial x} = 2\sum_i^{\text{DO}} h_{ii}^x + 2\sum_{ij}^{\text{DO}} \left\{ (ii|jj)^x - \frac{1}{2}(ij|ij)^x \right\} + 4\sum_i^{\text{MO}}\sum_j^{\text{DO}} U_{ij}^x F_{ij}. \quad (7.25)$$

If the molecular orbitals have been determined variationally, then the Fock matrix in the molecular orbital basis is diagonal, $F_{ij} = \epsilon_i \delta_{ij}$, which enables the further simplification

$$\frac{\partial E}{\partial x} = 2\sum_i^{\text{DO}} h_{ii}^x + 2\sum_{ij}^{\text{DO}} \left\{ (ii|jj)^x - \frac{1}{2}(ij|ij)^x \right\} + 4\sum_i^{\text{DO}} U_{ii}^x \epsilon_i. \quad (7.26)$$

As expected, the first derivative of the electronic energy includes core integral first derivatives (the first two sums) and molecular orbital coefficient first derivatives (the final sum).

Now comes the key simplification: it is actually possible to eliminate the first derivatives of the molecular orbital coefficients entirely by making use of the clever observation that

the molecular orbitals are orthonormal,

$$S_{ij} = \delta_{ij}, \qquad (7.27)$$

and hence

$$\frac{\partial S_{ij}}{\partial x} = 0. \qquad (7.28)$$

Substituting both of these results into equation (7.19) for the first derivative of the overlap integral yields the key relationship

$$0 = S_{ij}^x + U_{ji}^x + U_{ij}^x, \qquad (7.29)$$

from which it follows that the diagonal terms satisfy

$$U_{ii}^x = -\frac{1}{2}S_{ii}^x. \qquad (7.30)$$

The beauty of equation (7.30) is that it enables the elimination of the first derivative of the molecular orbital coefficients from equation (7.26), yielding

$$\frac{\partial E}{\partial x} = 2\sum_i^{\text{DO}} h_{ii}^x + 2\sum_{ij}^{\text{DO}} \left\{ (ii|jj)^x - \frac{1}{2}(ij|ij)^x \right\} - 2\sum_i^{\text{DO}} S_{ii}^x \epsilon_i. \qquad (7.31)$$

Hence, no explicit computation of the changes in the molecular orbitals is required to find the first derivative of the energy with respect to a nuclear displacement. This result is a specific illustration of the more general Hellman-Feynman theorem, which states that the $(2n+1)^{\text{st}}$ derivative of the energy requires up to the nth derivative of the molecular orbitals [77, 62].

Having completed the differentiation of the electronic energy in the MO basis, we can translate the result back into the AO basis to obtain the final expression [197]

$$\frac{\partial E}{\partial x} = \sum_{\mu\nu}^{\text{AO}} P_{\mu\nu}\frac{\partial h_{\mu\nu}}{\partial x} + \frac{1}{2}\sum_{\mu\nu\rho\sigma}^{\text{AO}} P_{\mu\nu}P_{\rho\sigma}\left\{\frac{\partial(\mu\nu|\rho\sigma)}{\partial x} - \frac{1}{2}\frac{\partial(\mu\rho|\nu\sigma)}{\partial x}\right\} - \sum_{\mu\nu}^{\text{AO}} W_{\mu\nu}\frac{\partial S_{\mu\nu}}{\partial x}, \quad (7.32)$$

where

$$W_{\mu\nu} = 2 \sum_{i}^{\text{DO}} C_{\mu}^{i} C_{\nu}^{i} \epsilon_i \tag{7.33}$$

is the *energy-weighted density matrix.*

7.6 Second Derivative of the Electronic Energy

There are two approaches to evaluating the second derivative of the electronic energy: we can begin from the first derivative equation (7.32) in the AO basis or the first derivative equation (7.31) in the MO basis. Because most quantum chemistry packages evaluate and store integrals in the AO basis, this time it proves easier to differentiate the AO equation (7.32) to obtain [141]

$$
\begin{aligned}
\frac{\partial^2 E}{\partial x \partial y} =& \sum_{\mu\nu}^{\text{AO}} P_{\mu\nu} \frac{\partial^2 h_{\mu\nu}}{\partial x \partial y} + \frac{1}{2} \sum_{\mu\nu\rho\sigma}^{\text{AO}} P_{\mu\nu} P_{\rho\sigma} \left\{ \frac{\partial^2 (\mu\nu|\rho\sigma)}{\partial x \partial y} - \frac{1}{2} \frac{\partial^2 (\mu\rho|\sigma\nu)}{\partial x \partial y} \right\} \\
& - \sum_{\mu\nu}^{\text{AO}} W_{\mu\nu} \frac{\partial^2 S_{\mu\nu}}{\partial x \partial y} + \sum_{\mu\nu}^{\text{AO}} \frac{\partial P_{\mu\nu}}{\partial y} \frac{\partial h_{\mu\nu}}{\partial x} \\
& + \sum_{\mu\nu\rho\sigma}^{\text{AO}} \frac{\partial P_{\mu\nu}}{\partial y} P_{\rho\sigma} \left\{ \frac{\partial (\mu\nu|\rho\sigma)}{\partial x} - \frac{1}{2} \frac{\partial (\mu\rho|\nu\sigma)}{\partial x} \right\} \\
& - \sum_{\mu\nu}^{\text{AO}} \frac{\partial W_{\mu\nu}}{\partial y} \frac{\partial S_{\mu\nu}}{\partial x}.
\end{aligned}
\tag{7.34}
$$

The first three terms in equation (7.34) involve only core integral derivatives, but the remaining terms involve first derivatives of the density matrix $P_{\mu\nu}$ and the energy-weighted density matrix $W_{\mu\nu}$, and these quantities in turn involve the first derivatives of the molecular orbital coefficients.

We begin by differentiating the density matrix in equation (7.10) to obtain

$$\frac{\partial P_{\mu\nu}}{\partial y} = 2 \sum_{i}^{\text{DO}} \left(\frac{\partial C_{\mu}^{i}}{\partial y} C_{\nu}^{i} + C_{\mu}^{i} \frac{\partial C_{\nu}^{i}}{\partial y} \right), \tag{7.35}$$

and using equation (7.15) for the derivatives of the molecular orbital coefficients gives rise to the expression

$$\frac{\partial P_{\mu\nu}}{\partial y} = 2 \sum_i^{\text{DO}} \sum_m^{\text{MO}} U_{mi}^y \left(C_\mu^m C_\nu^i + C_\mu^i C_\nu^m \right). \tag{7.36}$$

Unlike with the first derivative of the electron energy, this time we are interested in non-diagonal elements U_{mi}^y which cannot simply be eliminated in favor of the core overlap integral derivatives. Hence, to obtain the second derivative of the electron energy, we require knowledge of the first derivatives of the molecular orbital coefficients. This is another specific demonstration of the Hellman-Feynman theorem, which states that the $(2n+1)^{\text{st}}$ derivative of the energy requires up to the nth derivative of the molecular orbitals [77, 62].

To differentiate the energy-weighted density matrix, we rewrite the definition of $W_{\mu\nu}$ in equation (7.33) as

$$W_{\mu\nu} = 2 \sum_{ij}^{\text{DO}} C_\mu^i C_\nu^j F_{ij}. \tag{7.37}$$

Transforming the Fock matrix from the MO basis to the AO basis yields

$$W_{\mu\nu} = 2 \sum_{ij}^{\text{DO}} \sum_{\rho\sigma}^{\text{AO}} C_\mu^i C_\nu^j C_\rho^i C_\sigma^j F_{\rho\sigma}, \tag{7.38}$$

and summing over the doubly-occupied molecular orbitals gives the convenient expression

$$W_{\mu\nu} = \frac{1}{2} \sum_{\rho\sigma}^{\text{AO}} P_{\mu\rho} P_{\nu\sigma} F_{\rho\sigma}. \tag{7.39}$$

Differentiating this expression gives [141]

$$\frac{\partial W_{\mu\nu}}{\partial y} = \frac{1}{2} \sum_{\rho\sigma}^{\text{AO}} \left(\frac{\partial P_{\mu\rho}}{\partial y} P_{\nu\sigma} + P_{\mu\rho} \frac{\partial P_{\nu\sigma}}{\partial y} \right) F_{\rho\sigma} + \frac{1}{2} \sum_{\rho\sigma}^{\text{AO}} P_{\mu\rho} P_{\nu\sigma} \frac{\partial F_{\rho\sigma}}{\partial y}, \tag{7.40}$$

from which it is clear that we need the derivative of the first derivative of the Fock matrix with respect to nuclear displacements.

Beginning from the definition of the Fock matrix in the AO basis in equation (7.61) and taking the first derivative gives [141]

$$\frac{\partial F_{\mu\nu}}{\partial y} = F_{\mu\nu}^y + \sum_{\rho\sigma}^{AO} \frac{\partial P_{\rho\sigma}}{\partial y} \left\{ (\mu\nu|\rho\sigma) - \frac{1}{2}(\mu\rho|\sigma\nu) \right\},$$ (7.41)

where the first term

$$F_{\mu\nu}^y = \frac{\partial h_{\mu\nu}}{\partial y} + \sum_{\rho\sigma}^{AO} P_{\rho\sigma} \left\{ \frac{\partial(\mu\nu|\rho\sigma)}{\partial y} - \frac{1}{2}\frac{\partial(\mu\rho|\sigma\nu)}{\partial y} \right\}$$ (7.42)

is known as the *core* first derivative Fock matrix where all of the integrals that make up the Fock matrix have simply been replaced by their first derivatives, and the second term involves the first derivative of the density matrix, which in turn involves derivatives of the molecular orbital coefficients.

7.7 The Coupled-Perturbed Equations

The second derivative expression for the electronic energy in equation (7.34) involves the first derivative of the density matrix (7.36), which raises the question of how to compute the first derivatives of the molecular orbital coefficients which appear in the U_{ij}^y terms. Unlike with the first derivative of the electronic energy, where the derivatives of the molecular orbitals could be eliminated, this question can no longer be avoided. The key fact is that these molecular orbital coefficients, and hence their first derivatives, are not generic but are determined by variational optimization to give the lowest possible electronic energy.

In fact, we already made use of this fact when we assumed the Fock matrix was diagonal in the MO basis, $F_{ij} = \epsilon_i \delta_{ij}$, and the *derivative* of this condition gives rise to the *coupled-perturbed equations* whose solution yields the first derivative of the molecular orbital coefficients. Differentiating the condition $F_{ij} = \epsilon_i \delta_{ij}$ reveals that the derivative of the Fock

matrix is also diagonal in the MO basis

$$\frac{\partial F_{ij}}{\partial y} = \delta_{ij} \frac{\partial \epsilon_i}{\partial y}, \tag{7.43}$$

which implies that

$$\frac{\partial F_{ij}}{\partial y} = 0 \quad \text{for} \quad i \neq j. \tag{7.44}$$

It is this condition which gives rise to the *coupled-perturbed equations.*

To make use of this condition, we must compute the derivative of the Fock matrix in the MO basis, as given in equation (7.8), to obtain

$$\frac{\partial F_{ij}}{\partial y} = \frac{\partial h_{ij}}{\partial y} + \sum_{k}^{\text{DO}} \left\{ 2\frac{\partial (ij|kk)}{\partial y} - \frac{\partial (ik|ik)}{\partial y} \right\}. \tag{7.45}$$

Substituting the derivatives of the one-electron and two-electron integrals in equations (7.20) and (7.22) gives

$$\frac{\partial F_{ij}}{\partial y} = F_{ij}^y + \sum_{k}^{\text{MO}} \left(U_{ki}^y F_{kj} + U_{kj}^y F_{ik} \right) + \sum_{k}^{\text{MO}} \sum_{l}^{\text{DO}} U_{kl}^y A_{ij,kl}, \tag{7.46}$$

where

$$A_{ij,kl} = 4(ij|kl) - (ik|jl) - (il|jk) \tag{7.47}$$

is known as the *A*-matrix. Taking the Fock matrix to be diagonal so that $F_{ik} = \epsilon_i \delta_{ik}$ and $F_{kj} = \epsilon_j \delta_{jk}$ allows the implication

$$\frac{\partial F_{ij}}{\partial y} = F_{ij}^y + U_{ji}^y \epsilon_j + U_{ij}^y \epsilon_i + \sum_{k}^{\text{MO}} \sum_{l}^{\text{DO}} U_{kl}^y A_{ij,kl}. \tag{7.48}$$

Exploiting the orthonormality of the molecular orbitals in equation (7.29) implies that

$$U_{ji}^y = -S_{ij}^y - U_{ij}^y, \tag{7.49}$$

125

so that the derivative of the Fock matrix becomes

$$\frac{\partial F_{ij}}{\partial y} = F_{ij}^y - \left(\epsilon_j - \epsilon_i\right)U_{ij}^y - S_{ij}^y\epsilon_j + \sum_k^{DO}\sum_l^{DO} U_{kl}^y A_{ij,kl} + \sum_k^{VO}\sum_l^{DO} U_{kl}^y A_{ij,kl}, \qquad (7.50)$$

where the sum over all k molecular orbitals (MO) has been split into a sum over occupied orbitals (DO) and a sum over virtual orbitals (VO). By relabeling indices and making use of the orthonormality of the molecular orbitals in equation (7.29), as well as the definition of the A-matrix in equation 7.47, the fourth term may be rewritten as

$$\sum_{kl}^{DO} U_{kl}^y A_{ij,kl} = -\frac{1}{2}\sum_{kl}^{DO} S_{kl}^y\left\{4(ij|kl) - (ik|jl) - (il|jk)\right\} = -\sum_{kl}^{DO} S_{kl}^y\left\{2(ij|kl) - (ik|jl)\right\},$$

$$(7.51)$$

which yields the final expression for the derivative of the Fock matrix

$$\frac{\partial F_{ij}}{\partial y} = F_{ij}^y - \left(\epsilon_j - \epsilon_i\right)U_{ij}^y - S_{ij}^y\epsilon_j - \sum_{kl}^{DO} S_{kl}^y\left\{2(ij|kl) - (ik|j)\right\} + \sum_k^{VO}\sum_l^{DO} U_{kl}^y A_{ij,kl}. \quad (7.52)$$

Using the variational condition that the off-diagonal elements of this matrix are equal to zero, as per equation (7.44), yields the *coupled-perturbed equation* [141, 197]

$$\left(\epsilon_j - \epsilon_i\right)U_{ij}^y - \sum_y^{VO}\sum_l^{DO} U_{kl}^y A_{ij,kl} = B_{ij}^y \qquad (7.53)$$

where the B^y-vector is defined for each derivative along the y-direction as

$$B_{ij}^y = F_{ij}^y - S_{ij}^y\epsilon_j - \sum_{kl}^{DO} S_{kl}^y\left\{2(ij|kl) - (ik|jl)\right\}. \qquad (7.54)$$

For each derivative along the y-direction, equation (7.55) constitutes a system of coupled equations for the unknown variables U_{ij}^y, where i corresponds to a virtual molecular orbital and j corresponds to a doubly-occupied molecular orbital. Hence, the dimension of this system of coupled equations is the number of doubly-occupied molecular orbitals times the number of virtual molecular orbitals. Solving this system is typically the most

126

time-consuming step in evaluating the second derivative of the electronic energy, and it is typically accomplished by an iterative method [99].

After this system of coupled-perturbed equations has been solved, the remaining U_{ij}^y, where i and j both correspond to doubly-occupied molecular orbitals or both correspond to virtual molecular orbitals, may be found by simple rearrangement of equation (7.55),

$$U_{ij}^y = \frac{1}{\left(\epsilon_j - \epsilon_i\right)} \left(B_{ij}^y + \sum_k^{VO} \sum_l^{DO} U_{kl}^y A_{ij,kl} \right). \tag{7.55}$$

This can give rise to numerical instability if ϵ_i and ϵ_j are near-degenerate, and in practice computing these remaining U_{ij}^y can be avoided completely. It can be shown from the orthogonality relationship in equation (7.29) that when both i and j correspond to doubly-occupied molecular orbitals, we can substitute U_{ij}^y by $-\frac{1}{2} S_{ij}^y$, and when both i and j correspond to virtual molecular orbitals, we can substitute U_{ij}^y by 0 [141, 134].

7.8 Derivatives of the Nuclear-Nuclear Repulsion

The total energy of a molecule is the sum of its electronic energy and the energy arising from nuclear-nuclear repulsion,

$$E_{\text{nuc}} = \sum_{A>B}^{N} \frac{Z_A Z_B}{r_{AB}} \tag{7.56}$$

where Z_A is the charge of nucleus A, and r_{AB}, defined by

$$r_{AB}^2 = \left(x_A - x_B \right)^2 + \left(y_A - y_B \right)^2 + \left(z_A - z_B \right)^2, \tag{7.57}$$

is the distance between nuclei A and B.

The first derivative of the nuclear-nuclear repulsion energy is given by

$$\frac{\partial E_{\text{nuc}}}{\partial x_A} = - \sum_{B \neq A}^{N} \left(x_A - x_B \right) \frac{Z_A Z_B}{r_{AB}^3}. \tag{7.58}$$

There are four unique types of second derivatives, given by [197]

$$\frac{\partial^2 E_{\text{nuc}}}{\partial x_A^2} = -\sum_{B \neq A}^{N} \frac{Z_A Z_B}{r_{AB}^3} + 3 \sum_{B \neq A}^{N} \left(x_A - x_B \right)^2 \frac{Z_A Z_B}{r_{AB}^5}, \tag{7.59}$$

$$\frac{\partial^2 E_{\text{nuc}}}{\partial x_A \partial x_B} = \frac{Z_A Z_B}{r_{AB}^3} - 3 \left(x_A - x_B \right)^2 \frac{Z_A Z_B}{r_{AB}^5},$$

$$\frac{\partial^2 E_{\text{nuc}}}{\partial x_A \partial y_A} = 3 \sum_{B \neq A}^{N} \left(x_A - x_B \right) \left(y_A - y_B \right) \frac{Z_A Z_B}{r_{AB}^5},$$

$$\frac{\partial^2 E_{\text{nuc}}}{\partial x_A \partial y_B} = -3 \left(x_A - x_B \right) \left(y_A - y_B \right) \frac{Z_A Z_B}{r_{AB}^5}.$$

7.9 Protocol in Q-Chem for Computing Second Derivatives

Combining all of the results from the previous section, the general protocol for the computation of second derivatives in Q-Chem may be summarized in the following steps [161]. It is important to note at the outset that many computational optimizations, such as exploiting relevant symmetries for the evaluation of integrals, are possible to reduce the computational time and memory required, all of which are documented via comments within the code itself. However, this procedure outlines the key steps and identifies the functions in Q-Chem responsible for carrying out each step, which will suffice for our discussion of the integration of compressed sensing in the next section.

1. Compute all necessary integrals $h_{\mu\nu}$ and $(\mu\nu|\rho\sigma)$, their first derivatives $\dfrac{\partial S_{\mu\nu}}{\partial y}$, $\dfrac{\partial h_{\mu\nu}}{\partial y}$, and $\dfrac{\partial (\mu\nu|\rho\sigma)}{\partial y}$, and their second derivatives $\dfrac{\partial^2 S_{\mu\nu}}{\partial x \partial y}$, $\dfrac{\partial^2 h_{\mu\nu}}{\partial x \partial y}$, and $\dfrac{\partial^2 (\mu\nu|\rho\sigma)}{\partial x \partial y}$, as well as the density matrix

$$P_{\mu\nu} = 2 \sum_{i}^{\text{DO}} C_\mu^i C_\nu^i,$$

the Fock matrix

$$F_{\mu\nu} = h_{\mu\nu} + \sum_{\rho\sigma}^{\text{AO}} P_{\rho\sigma} \left\{ (\mu\nu|\rho\sigma) - \frac{1}{2} (\mu\rho|\nu\sigma) \right\}, \tag{7.60}$$

128

and the core first derivative Fock matrix

$$F_{\rho\sigma}^y = \frac{\partial h_{\rho\sigma}}{\partial y} + \sum_{\rho\sigma}^{\text{AO}} P_{\rho\sigma}\left\{\frac{\partial(\mu\nu|\rho\sigma)}{\partial y} - \frac{1}{2}\frac{\partial(\mu\rho|\sigma\nu)}{\partial y}\right\}.$$

2. Transform these integrals as necessary from the AO basis into the MO basis to set up the coupled-perturbed equations. In particular, assemble the transformed integrals into the A-matrix

$$A_{ij,kl} = 4(ij|kl) - (ik|jl) - (il|jk)$$

and the B^y-vectors

$$B_{ij}^y = F_{ij}^y - S_{ij}^y\epsilon_j - \sum_{kl}^{\text{DO}} S_{kl}^y\left\{2(ij|kl) - (ik|jl)\right\},$$

where

$$F_{ij}^y = h_{ij}^y + \sum_{k}^{\text{DO}}\left\{2(ij|kk)^y - (ik|jk)^y\right\}$$

is the core first derivative Fock matrix. In Q-Chem, these first two steps are accomplished by the `CPSCF_Lag` function.

3. Solve the coupled-perturbed equations

$$\left(\epsilon_j - \epsilon_i\right)U_{ij}^y - \sum_{k}^{\text{VO}}\sum_{l}^{\text{DO}} U_{kl}^y A_{ij,kl} = B_{ij}^y$$

for the first derivatives of the molecular orbital coefficients U_{ij}^y. In Q-Chem, this step is accomplished by the `IterZP` function.

4. Use the molecular orbital coefficients to compute the derivative of the density

$$\frac{\partial P_{\mu\nu}}{\partial y} = 2\sum_{i}^{\text{DO}}\sum_{m}^{\text{MO}} U_{mi}^y\left(C_\mu^m C_\nu^i + C_\mu^i C_\nu^m\right).$$

In Q-Chem, this step is accomplished by the `Get_Px` function.

5. Use the derivative of the density to compute the full derivative of the Fock matrix

$$\frac{\partial F_{\mu\nu}}{\partial y} = F_{\mu\nu}^{y} + \sum_{\rho\sigma}^{\text{AO}} \frac{\partial P_{\rho\sigma}}{\partial y} \left\{ (\mu\nu|\rho\sigma) - \frac{1}{2}(\mu\rho|\sigma\nu) \right\}.$$

In Q-Chem, this step is accomplished by the `Make_Fx` function.

6. Use the derivative of the Fock matrix to compute the derivative of the energy-weighted density matrix

$$\frac{\partial W_{\mu\nu}}{\partial y} = 2 \sum_{\rho\sigma}^{\text{AO}} \left(\frac{\partial P_{\mu\rho}}{\partial y} P_{\nu\sigma} + P_{\mu\rho} \frac{\partial P_{\nu\sigma}}{\partial y} \right) F_{\rho\sigma} + 2 \sum_{\rho\sigma}^{\text{AO}} P_{\mu\rho} P_{\nu\sigma} \frac{\partial F_{\rho\sigma}}{\partial y}.$$

In Q-Chem, this step is accomplished by the `Get_Wx_SCF` function.

7. Use the derivative of the energy-weighted density matrix, as well as the derivative of the density matrix, to compute the second derivative of the energy

$$\frac{\partial^2 E}{\partial x \partial y} = \sum_{\mu\nu}^{\text{AO}} P_{\mu\nu} \frac{\partial^2 h_{\mu\nu}}{\partial x \partial y} + \frac{1}{2} \sum_{\mu\nu\rho\sigma}^{\text{AO}} P_{\mu\nu} P_{\rho\sigma} \left\{ \frac{\partial^2 (\mu\nu|\rho\sigma)}{\partial x \partial y} - \frac{1}{2} \frac{\partial^2 (\mu\rho|\sigma\nu)}{\partial x \partial y} \right\}$$
$$- \sum_{\mu\nu}^{\text{AO}} W_{\mu\nu} \frac{\partial^2 S_{\mu\nu}}{\partial x \partial y} + \sum_{\mu\nu}^{\text{AO}} \frac{\partial P_{\mu\nu}}{\partial y} \frac{\partial h_{\mu\nu}}{\partial x}$$
$$+ \sum_{\mu\nu\rho\sigma}^{\text{AO}} \frac{\partial P_{\mu\nu}}{\partial y} P_{\rho\sigma} \left\{ \frac{\partial (\mu\nu|\rho\sigma)}{\partial x} - \frac{1}{2} \frac{\partial (\mu\rho|\nu\sigma)}{\partial x} \right\}$$
$$- \sum_{\mu\nu}^{\text{AO}} \frac{\partial W_{\mu\nu}}{\partial y} \frac{\partial S_{\mu\nu}}{\partial x} + \frac{\partial^2 E_{\text{nuc}}}{\partial x \partial y},$$

where the second derivative of the nuclear-nuclear repulsion energy is computed according to the appropriate formula

$$\frac{\partial^2 E_{\text{nuc}}}{\partial x_A^2} = -\sum_{B \neq A}^{N} \frac{Z_A Z_B}{r_{AB}^3} + 3 \sum_{B \neq A}^{N} \left(x_A - x_B \right)^2 \frac{Z_A Z_B}{r_{AB}^5},$$

$$\frac{\partial^2 E_{\text{nuc}}}{\partial x_A \partial y_B} = \frac{Z_A Z_B}{r_{AB}^3} - 3 \left(x_A - x_B \right)^2 \frac{Z_A Z_B}{r_{AB}^5},$$

$$\frac{\partial^2 E_{\text{nuc}}}{\partial x_A \partial y_A} = 3 \sum_{B \neq A}^{N} \left(x_A - x_B \right) \left(y_A - y_B \right) \frac{Z_A Z_B}{r_{AB}^5},$$

$$\frac{\partial^2 E_{\text{nuc}}}{\partial x_A \partial y_B} = -3 \left(x_A - x_B \right) \left(y_A - y_B \right) \frac{Z_A Z_B}{r_{AB}^5}.$$

In Q-Chem, this step is accomplished by the `SCF_Hess` function.

8. Mass-weight the matrix $\dfrac{\partial^2 E}{\partial x \partial y}$ and diagonalize it to obtain the vibrational normal modes (as the eigenvectors) and vibrational frequencies (as the square root of the eigenvalues).

7.10 Modifications in Q-Chem for Acceleration via Compressed Sensing

As discussed in the previous chapter and in ref. [154], the first requirement for applying compressed sensing to the computation of Hessian matrices is finding a basis in which the Hessian matrix is sparse. One choice is a basis of approximate normal modes, as may be computed by using a cheap molecular mechanics or electronic structure method, in which the more exact Hessian is expected to be nearly diagonal with a few small off-diagonal entries. These approximate normal modes are then transformed according to the discrete cosine transform (DCT). The Hessian is undersampled in this DCT basis, with only a few of the columns being computed, and compressed sensing is used to recover the complete sparse Hessian in the approximate normal mode basis. Once the complete Hessian has been recovered, it may be transformed back into the atomic coordinate basis and diagonalized to obtain the vibrational modes and frequencies.

As pertains to computing second derivatives of the energy, the key requirement for compressed sensing is that these derivatives must be computed not in the usual atomic basis but rather in the discrete cosine transform basis of the approximate normal modes. The advantage is that, in this DCT basis, only a small undersampling of the derivatives need to be computed [154]. In what follows, we use upper-case letters X and Y to denote

derivatives in the DCT basis, while we continue to use lower-case letters x and y to denote derivatives in the atomic nuclei basis.

The most computationally-expensive part of the Hessian matrix involves the terms that depend on the first derivative of the density matrix, $\dfrac{\partial P_{\mu\nu}}{\partial y}$, as these are the terms that require solving the coupled-perturbed equations for the derivatives of the molecular orbital coefficients [65, 99]. Table 7.1 confirms that solving the coupled-perturbed equations is the dominant part of a second derivative calculation, and the dominance increases as the molecule gets larger [65]. Hence, the key change to the second-derivative algorithm described above is that we solve the coupled-perturbed equations directly in the DCT basis, thus directly obtaining the first derivative of the density matrix, $\dfrac{\partial P_{\mu\nu}}{\partial Y}$, in the DCT basis. As only an undersampling of these derivatives is required, the key computational savings is that we no longer need to solve all $3N$ coupled-perturbed equations, where N is the number of atoms in the molecule, but instead only solve an undersampling of these equations.

The remainder of the Hessian matrix involves integrals over the atomic basis functions, as well as first and second derivatives of these integrals. Since the atomic basis functions are centered on individual atoms, it is advantageous to compute this part of the Hessian matrix directly in the atomic nuclei basis. This means that no undersampling is possible for this part of the Hessian matrix, as would be possible in the DCT basis. However, since this part of the Hessian matrix is less expensive to compute than the parts that depend on the first derivative of the density matrix, there is no computational bottleneck associated with computing it in the atomic nuclei basis in the conventional way without undersampling.

These changes result in the following modified procedure for computing the second derivatives of the energy, which can be toggled on and off by setting the value of the new Q-Chem keyword CS_HESS to 1 and 0 respectively:

Stage of Calculation	$C_4H_8N_2O_2$	$C_{14}H_{14}N_2O$
SCF	968	7468
Integral Derivatives	310	2568
Integral Second Derivatives	1005	10589
Coupled-Perturbed Equations	2392	51937
Total	5195	73175

Table 7.1: This table illustrates the time required for various stages of a second derivative calculation using the so-called "direct" method of analytic derivatives. The first calculation is α-formylaminopropanamide ($C_4H_8N_2O_2$) and the second is a cyan dye ($C_{14}H_{14}N_2O$). Each calculation was done at the Hartree-Fock level of theory with a 3-21G basis set [65] on a Multiflow TRACE 14/300 computer. All times are in seconds. It may be seen that solving the coupled-perturbed equations is the longest part of the calculation, and becomes more dominant as the number of atoms increases.

1. Compute all necessary integrals $h_{\mu\nu}$ and $(\mu\nu|\rho\sigma)$, their first derivatives $\dfrac{\partial S_{\mu\nu}}{\partial y}$, $\dfrac{\partial h_{\mu\nu}}{\partial y}$, and $\dfrac{\partial(\mu\nu|\rho\sigma)}{\partial y}$, and their second derivatives $\dfrac{\partial^2 S_{\mu\nu}}{\partial x \partial y}$, $\dfrac{\partial^2 h_{\mu\nu}}{\partial x \partial y}$, and $\dfrac{\partial^2(\mu\nu|\rho\sigma)}{\partial x \partial y}$, as well as the density matrix

$$P_{\mu\nu} = 2\sum_i^{\text{DO}} C_\mu^i C_\nu^i,$$

the Fock matrix

$$F_{\mu\nu} = h_{\mu\nu} + \sum_{\rho\sigma}^{\text{AO}} P_{\rho\sigma} \left\{ (\mu\nu|\rho\sigma) - \frac{1}{2}(\mu\rho|\nu\sigma) \right\}, \tag{7.61}$$

and the core first derivative Fock matrix

$$F_{\rho\sigma}^y = \frac{\partial h_{\rho\sigma}}{\partial y} + \sum_{\rho\sigma}^{\text{AO}} P_{\rho\sigma} \left\{ \frac{\partial(\mu\nu|\rho\sigma)}{\partial y} - \frac{1}{2}\frac{\partial(\mu\rho|\sigma\nu)}{\partial y} \right\},$$

where all the derivatives are in the atomic nuclei basis.

2. Compute the part of the first derivative of the energy-weighted density matrix that does not depend on the first derivative of the density matrix,

$$\frac{\partial W_{\mu\nu}^{(1)}}{\partial y} = 2\sum_{\rho\sigma}^{\text{AO}} P_{\mu\rho} P_{\nu\sigma} F_{\rho\sigma}^y,$$

where this derivative is in the atomic nuclei basis. In Q-Chem, this step is accomplished by the new `Get_Wx_SCF_CS1` function.

3. Use this result to compute all the contributions to the the second derivative of the energy that do not depend on the first derivative of the density matrix,

$$\frac{\partial^2 E^{(1)}}{\partial x \partial y} = \sum_{\mu\nu}^{\text{AO}} P_{\mu\nu} \frac{\partial^2 h_{\mu\nu}}{\partial x \partial y} + \frac{1}{2}\sum_{\mu\nu\rho\sigma}^{\text{AO}} P_{\mu\nu} P_{\rho\sigma} \left\{ \frac{\partial^2(\mu\nu|\rho\sigma)}{\partial x \partial y} - \frac{1}{2}\frac{\partial^2(\mu\rho|\sigma\nu)}{\partial x \partial y} \right\}$$
$$- \sum_{\mu\nu}^{\text{AO}} W_{\mu\nu} \frac{\partial^2 S_{\mu\nu}}{\partial x \partial y} - \sum_{\mu\nu}^{\text{AO}} \frac{\partial W_{\mu\nu}^{(1)}}{\partial y} \frac{\partial S_{\mu\nu}}{\partial x} + \frac{\partial^2 E_{\text{nuc}}}{\partial x \partial y},$$

where all the derivatives are in the atomic nuclei basis and the second derivative of the nuclear-nuclear repulsion energy is computed according to the usual formula (7.59). In Q-Chem, this step is now accomplished by the `SCF_Hess_CS1` function.

4. Transform the integrals as necessary from the AO basis into the MO basis to set up the coupled-perturbed equations. In particular, assemble the transformed integrals into the A-matrix

$$A_{ij,kl} = 4(ij|kl) - (ik|jl) - (il|jk)$$

and the B^y-vectors

$$B_{ij}^y = F_{ij}^y - S_{ij}^y \epsilon_j - \sum_{kl}^{\text{DO}} S_{kl}^y \left\{ 2(ij|kl) - (ik|jl) \right\},$$

where

$$F_{ij}^y = h_{ij}^y + \sum_{k}^{\text{DO}} \left\{ 2(ij|kk)^y - (ik|jk)^y \right\}$$

is the core first derivative Fock matrix. In Q-Chem, this step is still accomplished by the `CPSCF_Lag` function.

5. Transform the B_y-vectors, which involve derivatives in the basis of atomic nuclei, into the B_Y-vectors, which involve derivatives in the discrete cosine transform (DCT) basis described above. In Q-Chem, this step is accomplished by the new `Transform_Lag` function.

6. Solve an *undersampling* of the coupled-perturbed equations

$$\left(\epsilon_j - \epsilon_i \right) U_{ij}^Y - \sum_{k}^{\text{VO}} \sum_{l}^{\text{DO}} U_{kl}^Y A_{ij,kl} = B_{ij}^Y$$

for the first derivatives of the molecular orbital coefficients U_{ij}^Y. These derivatives are now obtained in the DCT basis, and the number of coupled-perturbed equations which are solved is equal to three times the value of the user-specified keyword `CS_ATOMS_PER_ITER` (up to a maximum of three times the total number of atoms in the molecule). In Q-Chem, the solving of these coupled-perturbed equations can still be done by the same `IterZP` function.

7. Use the molecular orbital coefficients to compute the derivative of the density in the DCT basis

$$\frac{\partial P_{\mu\nu}}{\partial Y} = 2 \sum_i^{DO} \sum_m^{MO} U_{mi}^Y \left(C_\mu^m C_\nu^i + C_\mu^i C_\nu^m \right).$$

Note that only an undersampling of density derivatives are computed, corresponding to the undersampling of coupled-perturbed equations which were solved. In Q-Chem, this step is still accomplished by the `Get_Px` function.

8. Use the derivative of the density to compute its contribution to the first derivative of the Fock matrix

$$\left(\frac{\partial F_{\mu\nu}}{\partial Y} \right)_{\text{no core}} = \sum_{\rho\sigma}^{AO} \frac{\partial P_{\rho\sigma}}{\partial Y} \left\{ (\mu\nu|\rho\sigma) - \frac{1}{2}(\mu\rho|\sigma\nu) \right\}.$$

This derivative is in the DCT basis and *excludes* the core Fock matrix derivative, which was accounted for previously. Note again that only an undersampling of derivatives are computed. In Q-Chem, this step is still accomplished by the `Make_Fx` function.

9. Use this result to compute the part of the derivative of the energy-weighted density matrix that depends on the derivative of the density matrix

$$\frac{\partial W_{\mu\nu}^{(2)}}{\partial Y} = 2 \sum_{\rho\sigma}^{AO} \left(\frac{\partial P_{\mu\rho}}{\partial Y} P_{\nu\sigma} + P_{\mu\rho} \frac{\partial P_{\nu\sigma}}{\partial Y} \right) F_{\rho\sigma} + 2 \sum_{\rho\sigma}^{AO} P_{\mu\rho} P_{\nu\sigma} \left(\frac{\partial F_{\mu\nu}}{\partial Y} \right)_{\text{no core}}.$$

Note again that only an undersampling of derivatives are computed. In Q-Chem, this step is accomplished by the new `Get_Wx_SCF_CS2` function.

10. Use this result to compute all the contributions to the second derivative of the energy that depend on the first derivative of the density matrix

$$
\frac{\partial^2 E^{(2)}}{\partial x \partial Y} = \sum_{\mu\nu}^{\text{AO}} \frac{\partial P_{\mu\nu}}{\partial Y} \frac{\partial h_{\mu\nu}}{\partial x}
$$
$$
+ \sum_{\mu\nu\rho\sigma}^{\text{AO}} \frac{\partial P_{\mu\nu}}{\partial Y} P_{\rho\sigma} \left\{ \frac{\partial(\mu\nu|\rho\sigma)}{\partial x} - \frac{1}{2} \frac{\partial(\mu\rho|\nu\sigma)}{\partial x} \right\}
$$
$$
- \sum_{\mu\nu}^{\text{AO}} \frac{\partial W_{\mu\nu}^{(2)}}{\partial Y} \frac{\partial S_{\mu\nu}}{\partial x}.
$$

In Q-Chem, this step is accomplished by the new `SCF_Hess_CS2` function. Note that only an undersampling of Y derivatives have been computed in the DCT basis, which corresponds to having only certain *columns* of the second derivative matrix $\frac{\partial^2 E^{(2)}}{\partial x \partial Y}$ available. The number of available columns is equal to the number of coupled-perturbed equations which were solved.

11. At this point, the entire matrix $\frac{\partial^2 E^{(1)}}{\partial x \partial y}$, which are the contributions to the second derivative that do *not* depend on the derivatives of the molecular orbital coefficients, has been computed in the atomic nuclei basis. In addition, certain *columns* of the matrix $\frac{\partial^2 E^{(2)}}{\partial x \partial Y}$, which are the contributions of the second derivative that *do* depend on the derivatives of the molecular orbital coefficients, have been computed with the Y derivative in the DCT basis. Convert both of these quantities fully into the DCT basis and add them together,

$$
\frac{\partial^2 E}{\partial X \partial Y} = \frac{\partial^2 E^{(1)}}{\partial X \partial Y} + \frac{\partial^2 E^{(2)}}{\partial X \partial Y},
$$

to obtain an undersampling of the *columns* of the full second derivative matrix in the DCT basis. In Q-Chem, this step is also done in the `SCF_Hess_CS2` function.

12. Apply compressed sensing, as fully described in the previous chapter and in ref. [154], to attempt to recover the full second-derivative matrix in the sparse basis of approximate normal modes. Then repeat the compressed sensing procedure using one less column of the matrix in step 11 to see whether the results are converged. The results are considered converged when the relative Frobenius norm error between the two recovered matrices is less than $10^{-\texttt{CS\_TOL}}$, where $\texttt{CS\_TOL}$ is a user-specified keyword. In Q-Chem, this compressed sensing procedure is done by the new function $\texttt{bpdn}$ which solves the basis-pursuit denoising problem using the algorithm described in ref. [184].

13. If the results are not converged, go back to Step 6 and compute more Y derivatives of the density matrix $\dfrac{\partial P_{\mu\nu}}{\partial Y}$ in order to obtain more columns of the second-derivative matrix $\dfrac{\partial^2 E}{\partial X \partial Y}$. If the results are converged, convert the recovered matrix back into the atomic nuclei basis $\dfrac{\partial^2 E}{\partial x \partial y}$, mass-weight the matrix, and diagonalize it to obtain the vibrational normal modes (as the eigenvectors) and vibrational frequencies (as the square root of the eigenvalues).

To summarize, the key advantage of compressed sensing is that only an undersampling of the coupled-perturbed equations are solved to yield the first derivative of the density matrix, $\dfrac{\partial P_{\mu\nu}}{\partial Y}$, directly in the DCT basis. As solving the coupled-perturbed equations is typically the most expensive part of a second derivative computation, this undersampling can result in substantial computational savings. The remaining parts of the Hessian, which do not depend on the first derivative of the density matrix, are still computed directly in the atomic basis, and subsequently transformed to the DCT basis, after which the compressed sensing procedure is applied.

So far, we have not observed a reduction in CPU hours when analytical second derivatives are computed with compressed sensing. While accurate vibrational normal modes and

frequencies are obtained by solving fewer coupled-perturbed equations, this does not appear to translate into fewer CPU hours. Even though fewer coupled-perturbed equations must be solved with compressed sensing, the total time spent solving these equations is comparable to the traditional approach. Understanding why solving fewer coupled-perturbed equations does not translate into a computational speed-up is the subject of ongoing research. Nevertheless, the method above is presented in the hopes that it might prove useful in other electronic software packages, with different approaches taken to solve the coupled-perturbed equations, and that we might also come to understand how to achieve the desired reduction in CPU hours in Q-Chem.

Chapter 8

Conclusion

Compressed sensing is revolutionizing how information is measured or computed throughout the physical sciences and beyond. Many systems have far less information than meets the eye; they might be governed, for example, by a few key parameters. Yet extracting that information, if one does not already know it in the first place, is a challenging task. The typical approach, colloquially termed "collect then compress," is to measure or compute a large amount of data, and then to compress that data into a better representation that reveals the true information content. While at first glance this might seem like the only possible generic approach, the development of compressed sensing has revealed that this approach is far from optimal. Instead, compressed sensing provides a general mathematical framework that makes it possible to measure or compute an amount of data that is *proportional* to the amount of information present in the data.

In this thesis, we have applied the compressed sensing paradigm to two problems in chemistry: one in the area of spectroscopy and one in the area of quantum chemistry and electronic structure. In traditional approaches to spectroscopy, it is necessary to acquire an entire spectrum while all of the information content is present in the peaks. In contrast, we have shown that compressed sensing makes it possible to acquire a spectrum with a

much smaller number of measurements proportional to the number of peaks, provided the measurements are taken in such a way that all of the peaks contribute to all of the measurements, so that each measurement is maximizing the amount of new information being obtained about the system. We have applied the methods to the computation of vibrational, electronic absorption, and circular dichroism spectra in the framework of real-time time-dependent density functional theory, the experimental measurement of multidimensional atomic spectra, and the extraction of spectral densities from molecular dynamics simulations.

In traditional approaches to quantum chemistry, an entire Hessian matrix is computed using an expensive electronic structure method to obtain accurate molecular vibrations and their frequencies, yet the expensive method provides only a limited amount of "corrective" information over a cheaper method. Using compressed sensing, we have shown that it is possible to compute accurate vibrational modes and frequencies with a smaller number of computations that is proportional only to the amount of *new* information the expensive method provides over the cheaper method.

The general lesson of this thesis is to think carefully about where genuinely new chemical information is found within a sea of chemical data. By redesigning the measurement process, and applying compressed sensing, it becomes possible to obtain this new chemical information without first measuring or computing the entire sea of data. As more people begin working at the interface of chemistry, applied math, and information theory, we expect that compressed sensing will find many more applications throughout chemistry.

In fact, several other chemical applications have sprung up in other sub-fields during the creation of this thesis. Compressed sensing has become a tool not only in vibrational and electronic spectroscopy, but also in nuclear magnetic resonance (NMR) spectroscopy [81].

At the same time, ideas from compressed sensing have been used to find compact wavefunction representations to describe many-body quantum systems, such as molecules, without detailed knowledge of the system. Applying the technique on the ground-state of hydrogen fluoride within the full configuration interaction, it is possible to obtain 98% of the correlation energy using only 50 configurations out of a possible 10^7 [119]. In this NOMAGIC method, the ideas of compressed sensing are harnessed to find those 50 configurations that best capture the correlation energy. In the right representation, there is far less correlation information than first meets the eye.

Compressed sensing has been described as a change from "measure what *can* be measured" to "measure what *should* be measured" [172]. This thesis, along with other applications beginning to emerge in other chemical sub-fields, represents a change from "compute what *can* be computed" to "compute what *should* be computed."

Bibliography

[1] D. Abramavicius and S. Mukamel. Quantum oscillatory exciton migration in photosynthetic reaction centers. *J. Chem. Phys.*, **133**, 064510 (2010).

[2] J. Adolphs and T. Renger. How Proteins Trigger Excitation Energy Transfer in the FMO Complex of Green Sulfur Bacteria. *Biophys. J.*, **91**, 2778 (2006).

[3] M. Aharon, M. Elad, and A. Bruckstein. K-SVD: An Algorithm for Designing Overcomplete Dictionaries for Sparse Representation. *IEEE Trans. Signal Process.*, **54**, 4311 (2006).

[4] D. Ahn. Time-convolutionless reduced-density-operator theory of an arbitrary driven system coupled to a stochastic reservoir: Quantum kinetic equations for semiconductors. *Phys. Rev. B*, **50**, 8310 (1994).

[5] M. P. Allen and D. J. Tildesley. *Computer Simulation of Liquids*. Oxford Science Publications. Clarendon Press, 1989.

[6] N. L. Allinger, Y. H. Yuh, and J. H. Lii. Molecular Mechanics. The MM3 Force Field for Hydrocarbons. 1. *J. Am. Chem. Soc.*, **111**, 8551 (1989).

[7] J. Almeida, J. Prior, and M. B. Plenio. Computation of Two-Dimensional Spectra Assisted by Compressed Sampling. *J. Phys. Chem. Lett.*, **3**, 2692 (2012).

[8] J. L. Alonso, X. Andrade, P. Echenique, F. Falceto, D. Prada-Gracia, and A. Rubio. Efficient Formalism for Large-Scale Ab Initio Molecular Dynamics based on Time-Dependent Density Functional Theory. *Phys. Rev. Lett.*, **101**, 096403 (2008).

[9] M. AlQuraishi and H. H. McAdams. Direct inference of protein–DNA interactions using compressed sensing methods. *Proc. Natl. Acad. Sci. U.S.A.*, **108**, 14819 (2011).

[10] X. Andrade, J. Alberdi-Rodriguez, D. A. Strubbe, M. J. T. Oliveira, F. Nogueira, A. Castro, J. Muguerza, A. Arruabarrena, S. G. Louie, A. Aspuru-Guzik, A. Rubio, and M. A. L. Marques. Time-dependent density-functional theory in massively parallel computer architectures: the octopus project. *J. Phys.: Condens. Matter*, **24**, 233202 (2012).

[11] X. Andrade, S. Botti, M. A. L. Marques, and A. Rubio. Time-dependent density functional theory scheme for efficient calculations of dynamic (hyper)polarizabilities. *J. Chem. Phys.*, **126**, 184106 (2007).

[12] X. Andrade, A. Castro, D. Zueco, J. L. Alonso, P. Echenique, F. Falceto, and A. Rubio. Modified Ehrenfest Formalism for Efficient Large-Scale ab initio Molecular Dynamics. *J. Chem. Theo. Comput.*, **5**, 728 (2009).

[13] X. Andrade, J. N. Sanders, and A. Aspuru-Guzik. Application of compressed sensing to the simulation of atomic systems. *Proc. Natl. Acad. Sci. U.S.A.*, **109**, 13928 (2012).

[14] Y. August and A. Stern. Compressive sensing spectrometry based on liquid crystal devices. *Opt. Lett.*, **38**, 4996 (2013).

[15] C. H. Baldwin, A. Kalev, and I. H. Deutsch. Quantum process tomography of unitary and near-unitary maps. *Phys. Rev. A*, **90**, 012110 (2014).

[16] R. G. Baraniuk. Compressive Sensing. *IEEE Signal Process. Mag.*, **24**, 118 (2007).

[17] R. G. Baraniuk, E. Candes, M. Elad, and Y. Ma. Applications of Sparse Representation and Compressive Sensing. *Proc. IEEE*, **98**, 906 (2010).

[18] S. Baroni, S. de Gironcoli, A. Dal Corso, and P. Giannozzi. Phonons and related crystal properties from density-functional perturbation theory. *Rev. Mod. Phys.*, **73**, 515 (2001).

[19] A. D. Becke. Density-functional exchange-energy approximation with correct asymptotic behavior. *Phys. Rev. A*, **38**, 3098 (1988).

[20] A. D. Becke. Density-functional thermochemistry. III. The role of exact exchange. *J. Chem. Phys.*, **98**, 5648 (1993).

[21] M. Beer and C. Ochsenfeld. Efficient linear-scaling calculation of response properties: Density matrix-based laplace-transformed coupled-perturbed self-consistent field theory. *J. Chem. Phys.*, **128**, 221102 (2008).

[22] P. H. Berens, S. R. White, and K. R. Wilson. Molecular dynamics and spectra. II. Diatomic Raman. *J. Chem. Phys.*, **75**, 515 (1981).

[23] T. C. Berkelbach, M. S. Hybertsen, and D. R. Reichman. Microscopic theory of singlet exciton fission. I. General formulation. *J. Chem. Phys.*, **138**, 114102 (2013).

[24] T. C. Berkelbach, M. S. Hybertsen, and D. R. Reichman. Microscopic theory of singlet exciton fission. II. Application to pentacene dimers and the role of superexchange. *J. Chem. Phys.*, **138**, 114103 (2013).

[25] G. F. Bertsch, J.-I. Iwata, A. Rubio, and K. Yabana. Real-space, real-time method for the dielectric function. *Phys. Rev. B*, **62**, 7998 (2000).

[26] J. M. Bioucas-Dias and M. A. T. Figueiredo. A New TwIST: Two-Step Iterative Shrinkage/Thresholding Algorithms for Image Restoration. *IEEE Trans. Image Process.*, **16**, 2992 (2007).

[27] J. M. Bioucas-Dias and M. A. T. Figueiredo. Two-Step Algorithms for Linear Inverse Problems with Non-Quadratic Regularization. *Proc. IEEE Int. Conf. Image Process.*, **1**, 105 (2007).

[28] H.-P. Breuer and F. Petruccione. *The Theory of Open Quantum Systems.* Oxford University Press, 2007.

[29] T. Brixner, T. Mancal, I. V. Stiopkin, and G. R. Fleming. Phase-stabilized two-dimensional electronic spectroscopy. *J. Chem. Phys.*, **121**, 4221 (2004).

[30] T. Brixner, J. Stenger, H. M. Vaswani, M. Cho, R. E. Blankenship, and G. R. Fleming. Two-dimensional spectroscopy of electronic couplings in photosynthesis. *Nature*, **434**, 625 (2005).

[31] E. Candès and J. Romberg. Sparsity and Incoherence in Compressive Sampling. *Inverse Prob.*, **23**, 969 (2007).

[32] E. J. Candès and C. Fernandez-Granda. Towards a Mathematical Theory of Super-resolution. *Comm. Pure Appl. Math.*, **67**, 906 (2014).

[33] E. J. Candès, X. Li, Y. Ma, and J. Wright. Robust principal component analysis? *J. ACM*, **58**, 11 (2011).

[34] E. J. Candès, J. Romberg, and T. Tao. Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information. *IEEE Trans. Inf. Theory*, **52**, 489 (2006).

[35] E. J. Candès and M. B. Wakin. An Introduction to Compressive Sampling. *IEEE Signal Process. Mag.*, **25**, 21 (2008).

[36] J. Cao and R. J. Silbey. Optimization of Exciton Trapping in Energy Transfer Processes. *J. Phys. Chem. A*, **113**, 13825 (2009).

[37] M. Carnell, S. D. Peyerimhoff, A. Breest, K. H. Gödderz, P. Ochmann, and J. Hormes. Experimental and quantum-theoretical investigation of the circular dichroism spectrum of R-methyloxirane. *Chem. Phys. Lett.*, **180**, 477 (1991).

[38] F. Caruso, S. K. Saikin, E. Solano, S. F. Huelga, A. Aspuru-Guzik, and M. B. Plenio. Probing biological light-harvesting phenomena by optical cavities. *Phys. Rev. B*, **85**, 125424 (2012).

[39] M. E. Casida. Time-dependent density functional response theory of molecular systems: theory, computational methods, and functionals. In J. M. Seminario, editor, *Recent Developments and Applications of Modern Density Functional Theory*, pages 391–440. Elsevier Science, 1996.

[40] A. Castro, H. Appel, M. Oliveira, C. A. Rozzi, X. Andrade, F. Lorenzen, M. A. L. Marques, E. K. U. Gross, and A. Rubio. octopus: a tool for the application of time-dependent density functional theory. *Phys. Status Solidi B*, **243**, 2465 (2006).

[41] A. Castro, M. A. L. Marques, J. A. Alonso, G. F. Bertsch, and A. Rubio. Excited states dynamics in time-dependent density functional theory. *Eur. Phys. J. D*, **28**, 211 (2004).

[42] A. Castro, M. A. L. Marques, J. A. Alonso, and A. Rubio. Optical properties of nanostructures from time-dependent density functional theory. *J. Comput. Theor. Nanosci.*, **1**, 231 (2004).

[43] M. Ceccarelli, P. Procacci, and M. Marchi. An ab initio force field for the cofactors of bacterial photosynthesis. *J. Comput. Chem.*, **24**, 129 (2003).

[44] M. Ceotto, S. Atahan, G. F. Tantardini, and A. Aspuru-Guzik. Multiple coherent states for first-principles semiclassical initial value representation molecular dynamics. *J. Chem. Phys.*, **130**, 234113 (2009).

[45] R. Chartrand, R. G. Baraniuk, Y. C. Eldar, M. A. T. Figueiredo, and J. Tanner. Introduction to the Issue on Compressive Sensing. *IEEE J. Sel. Topics Signal Process.*, **4**, 241 (2010).

[46] S. S. Chen, D. L. Donoho, and M. A. Saunders. Atomic Decomposition by Basis Pursuit. *SIAM Rev.*, **43**, 129 (2001).

[47] X. Chen, J. Cao, and R. J. Silbey. A novel construction of complex-valued Gaussian processes with arbitrary spectral densities and its application to excitation energy transfer. *J. Chem. Phys.*, **148**, 224104 (2013).

[48] A. W. Chin, J. Prior, R. Rosenbach, F. Caycedo-Soler, S. F. Huelga, and M. B. Plenio. The role of non-equilibrium vibrational structures in electronic coherence and recoherence in pigment-protein complexes. *Nature Phys.*, **9**, 113 (2013).

[49] N. Christensson, H. F. Kauffmann, T. Pullerits, and T. Mančal. Origin of Long-Lived Coherences in Light-Harvesting Complexes. *J. Phys. Chem. B*, **116**, 7449 (2012).

[50] W. D. Cornell, P. Cieplak, C. I. Bayly, I. R. Gould, K. M. Merz, D. M. Ferguson, D. C. Spellmeyer, T. Fox, J. W. Caldwell, and P. A. Kollman. A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules. *J. Am. Chem. Soc.*, **117**, 5179 (1995).

[51] W. K. Coulter, C. J. Hillar, G. Isley, and F. T. Sommer. Adaptive compressed sensing– a new class of self-organizing coding models for neuroscience. *Proc. IEEE Int. Conf. Acoust. Speech Signal Process.*, **35**, 5494 (2010).

[52] C. J. Cramer. *Essentials of Computational Chemistry: Theories and Models.* Wiley, 2nd edition, 2004.

[53] L. A. Curtiss, K. Raghavachari, P. C. Redfern, and J. A. Pople. Assessment of Gaussian-2 and density functional theories for the computation of enthalpies of formation. *J. Chem. Phys.*, **106**, 1063 (1997).

[54] P. Deglmann, F. Furche, and R. Ahlrichs. An efficient implementation of second analytical derivatives for density functional methods. *Chem. Phys. Lett.*, **362**, 511 (2002).

[55] J. M. Dickey and A. Paskin. Computer Simulation of the Lattice Dynamics of Solids. *Phys. Rev.*, **188**, 1407 (1969).

[56] M. Doneva, P. Börnert, H. Eggers, A. Mertins, J. Pauly, and M. Lustig. Compressed sensing for chemical shift-based water-fat separation. *Magn. Reson. Med.*, **64**, 1749 (2010).

[57] D. L. Donoho. Compressed sensing. *IEEE Trans. Inf. Theory*, **52**, 1289 (2006).

[58] D. L. Donoho and X. Huo. Uncertainty principles and ideal atomic decomposition. *IEEE Trans. Inf. Theory*, **47**, 2845 (2001).

[59] M. F. Duarte, M. A. Davenport, D. Takhar, J. N. Laska, T. Sun, K. F. Kelly, and R. G. Baraniuk. Single-Pixel Imaging via Compressive Sampling. *IEEE Signal Process. Mag.*, **25**, 83 (2008).

[60] M. Elad and A. Feuer. Restoration of a single superresolution image from several blurred, noisy, and undersampled measured images. *IEEE Trans. Image Process.*, **6**, 1646 (1997).

[61] R. R. Ernst and W. A. Anderson. Application of Fourier Transform Spectroscopy to Magnetic Resonance. *Rev. Sci. Instrum.*, **37**, 93 (1966).

[62] R. P. Feynman. Forces in Molecules. *Phys. Rev.*, **56**, 340 (1939).

[63] S. T. Flammia, D. Gross, Y.-K. Liu, and J. Eisert. Quantum tomography via compressed sensing: error bounds, sample complexity and efficient estimators. *New J. Phys.*, **14**, 095022 (2012).

[64] W. T. Freeman, T. R. Jones, and E. C. Pasztor. Example-based super-resolution. *IEEE Comput. Graph. Appl.*, **22**, 56 (2002).

[65] M. Frisch, M. Head-Gordon, and J. Pople. Direct analytic SCF second derivatives and electric field properties. *Chem. Phys.*, **141**, 189 (1990).

[66] J. Gerratt and I. M. Mills. Force Constants and Dipole-Moment Derivatives of Molecules from Perturbed Hartree-Fock Calculations. I. *J. Chem. Phys.*, **49**, 1719 (1979).

[67] A. Ghosh, J. Qiu, W. F. DeGrado, and R. M. Hochstrasser. Tidal surge in the M2 proton channel, sensed by 2D IR spectroscopy. *Proc. Natl. Acad. Sci. U.S.A.*, **108**, 6115 (2011).

[68] A. Ghysels, H. L. Woodcock, J. D. Larkin, B. T. Miller, Y. Shao, J. Kong, D. Van Neck, V. Van Speybroeck, M. Waroquier, and B. R. Brooks. Efficient Calculation of QM/MM Frequencies with the Mobile Block Hessian. *J. Chem. Theory Comput.*, **7**, 496 (2011).

[69] D. Gross, Y.-K. Liu, S. T. Flammia, S. Becker, and J. Eisert. Quantum State Tomography via Compressed Sensing. *Phys. Rev. Lett.*, **105**, 150401 (2010).

[70] S. F. Gull and G. J. Daniell. Image reconstruction from incomplete and noisy data. *Nature*, **272**, 686 (1978).

[71] P. Hamm and R. M. Hochstrasser. Structure and dynamics of proteins and peptides: Femtosecond two-dimensional infrared spectroscopy. In M. D. Fayer, editor, *Ultrafast Infrared and Raman Spectroscopy*, pages 273–347. CRC Press, 2001.

[72] N. C. Handy, J. F. Gaw, and E. D. Simandiras. Accurate ab initio prediction of molecular geometries and spectroscopic constants, using SCF and MP2 energy derivatives. *J. Chem. Soc. Faraday Trans.*, **83**, 1577 (1987).

[73] N. C. Handy and H. F. Schaefer. On the evaluation of analytic energy derivatives for correlated wave functions. *J. Chem. Phys.*, **81**, 5031 (1984).

[74] N. C. Handy, D. J. Tozer, G. J. Laming, C. W. Murray, and R. D. Amos. Analytic Second Derivatives of the Potential Energy Surface. *Isr. J. Chem.*, **33**, 331 (1993).

[75] B. Hein, C. Kreisbeck, T. Kramer, and M. Rodríguez. Modelling of oscillations in two-dimensional echo-spectra of the Fenna–Matthews–Olson complex. *New J. Phys.*, **14**, 023018 (2012).

[76] B. Heinz-Peter, B. Kappler, and F. Petruccione. Time-Convolutionless Stochastic Unraveling of Non-Markovian Quantum Master Equations. In P. Blanchard, D. Giulini, E. Joos, C. Kiefer, and I.-O. Stamatescu, editors, *Decoherence: Theoretical, Experimental, and Conceptual Problems*, pages 233–241. Springer, 2000.

[77] H. Hellmann. *Einführung in die Quantenchemie.* Leipzig und Wien: F. Deuticke, 1937.

[78] M. A. Herman and T. Strohmer. High-Resolution Radar via Compressed Sensing. *IEEE Trans. Signal Process.*, **57**, 2275 (2009).

[79] R. M. Hochstrasser. Multidimensional ultrafast spectroscopy. *Proc. Natl. Acad. Sci. U.S.A.*, **104**, 14189 (2007).

[80] P. Hohenberg and W. Kohn. Inhomogeneous Electron Gas. *Phys. Rev.*, **136**, B864 (1964).

[81] D. J. Holland, M. J. Bostock, L. F. Gladden, and D. Nietlispach. Fast Multidimensional NMR Spectroscopy Using Compressed Sensing. *Angew. Chem. Int. Ed.*, **50**, 6548 (2011).

[82] T. Hu, A. Leonardo, and D. B. Chklovskii. Reconstruction of Sparse Circuits Using Multi-neuronal Excitation (RESCUME). *Adv. Neural Inf. Process. Syst.*, **22**, 790 (2009).

[83] A. Ishizaki and G. R. Fleming. Theoretical examination of quantum coherence in a photosynthetic system at physiological temperature. *Proc. Natl. Acad. Sci. U.S.A.*, **106**, 17255 (2009).

[84] A. Ishizaki and G. R. Fleming. Unified treatment of quantum coherent and incoherent hopping dynamics in electronic energy transfer: Reduced hierarchy equation approach. *J. Chem. Phys.*, **130**, 234111 (2009).

[85] A. Ishizaki and G. R. Fleming. On the Interpretation of Quantum Coherent Beats Observed in Two-Dimensional Electronic Spectra of Photosynthetic Light Harvesting Complexes. *J. Phys. Chem. B*, **115**, 6227 (2011).

[86] S. Jang. Theory of multichromophoric coherent resonance energy transfer: A polaronic quantum master equation approach. *J. Chem. Phys.*, **135**, 034105 (2011).

[87] S. Jang, M. D. Newton, and R. J. Silbey. Multichromophoric Förster Resonance Energy Transfer from B800 to B850 in the Light Harvesting Complex 2: Evidence for Subtle Energetic Optimization by Purple Bacteria. *J. Phys. Chem. B*, **111**, 6807 (2007).

[88] D. M. Jonas. Two-Dimensional Femtosecond Spectroscopy. *Annu. Rev. Phys. Chem.*, **54**, 425 (2003).

[89] P. Jørgensen and J. Simons. Ab initio analytical molecular gradients and Hessians. *J. Chem. Phys.*, **79**, 334 (1983).

[90] E. S. Kadantsev and M. J. Stott. Calculation of vibrational frequencies within the real space pseudopotential approach. *Phys. Rev. B*, **71**, 045104 (2005).

[91] O. Katz, J. M. Levitt, and Y. Silberberg. Compressive Fourier Transform Spectroscopy. *Frontiers in Optics*, page FTuE3 (2010).

[92] K. Kazimierczuk and V. Y. Orekhov. Accelerated NMR Spectroscopy by Using Compressed Sensing. *Angew. Chem. Int. Ed.*, **50**, 5556 (2011).

[93] M. Khalil, N. Demirdöven, and A. Tokmakoff. Coherent 2D IR Spectroscopy: Molecular Structure and Dynamics in Solution. *J. Phys. Chem. A*, **107**, 5258 (2003).

[94] U. Kleinekathöfer. Non-Markovian theories based on a decomposition of the spectral density. *J. Chem. Phys.*, **121**, 2505 (2004).

[95] E. E. Koch and A. Otto. Optical absorption of benzene vapour for photon energies from 6 eV to 35 eV. *Chem. Phys. Lett.*, **12**, 476 (1972).

[96] P. Koehl. Linear Prediction Spectral Analysis of NMR Data. *Progr. Nucl. Magn. Reson. Spectrosc.*, **34**, 257 (1999).

[97] W. Kohn and L. J. Sham. Self-Consistent Equations Including Exchange and Correlation Effects. *Phys. Rev.*, **140**, A1133 (1965).

[98] A. Kolli, A. Nazir, and A. Olaya-Castro. Electronic excitation dynamics in multichromophoric systems described via a polaron-representation master equation. *J. Chem. Phys.*, **135**, 154112 (2011).

[99] P. P. Korambath, J. Kong, T. R. Furlani, and M. Head-Gordon. Parallelization of analytical Hartree-Fock and density functional theory Hessian calculations. Part I: parallelization of coupled-perturbed Hartree-Fock equations. *Mol. Phys.*, **100**, 1755 (2002).

[100] D. Kouame and M. Ploquin. Super-resolution in medical imaging: An illustrative approach through ultrasound. *Proc. IEEE Int. Symp. Biomed. Imaging*, **6**, 249 (2009).

[101] C. Kreisbeck and T. Kramer. Long-Lived Electronic Coherence in Dissipative Exciton Dynamics of Light-Harvesting Complexes. *J. Phys. Chem. Lett.*, **3**, 2828 (2012).

[102] C. Kreisbeck, T. Kramer, M. Rodríguez, and B. Hein. High-Performance Solution of Hierarchical Equations of Motion for Studying Energy Transfer in Light-Harvesting Complexes. *J. Chem. Theory Comput.*, **7**, 2166 (2011).

[103] R. Krishnan, H. B. Schlegel, and J. A. Pople. Derivative studies in configuration-interaction theory. *J. Chem. Phys.*, **72**, 4654 (1980).

[104] F. Krzakala, M. Mézard, F. Sausset, Y. Sun, and L. Zdeborová. Probabilistic reconstruction in compressed sensing: algorithms, phase diagrams, and threshold achieving matrices. *J. Stat. Mech.*, **2012**, P08009 (2012).

[105] F. Krzakala, M. Mézard, F. Sausset, Y. F. Sun, and L. Zdeborová. Statistical-Physics-Based Reconstruction in Compressed Sensing. *Phys. Rev. X*, **2**, 021005 (2012).

[106] J. Kussmann, M. Beer, and C. Ochsenfeld. Linear-scaling self-consistent field methods for large molecules. *WIREs Comput. Mol. Sci.*, **3**, 614 (2013).

[107] J. J. Led and H. Gesmar. Application of the Linear Prediction Method to NMR Spectroscopy. *Chem. Rev.*, **91**, 1413 (1991).

[108] C. Lee, W. Yang, and R. G. Parr. Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density. *Phys. Rev. B*, **37**, 785 (1988).

[109] K.-M. Lee, K. Yabana, and G. F. Bertsch. Magnetic circular dichroism in real-time time-dependent density functional theory. *J. Chem. Phys.*, **134**, 144106 (2011).

[110] G. A. Lott, A. Perdomo-Ortiz, J. K. Utterback, J. R. Widom, A. Aspuru-Guzik, and A. H. Marcus. Conformation of self-assembled porphyrin dimers in liposome vesicles by phase-modulation 2D fluorescence spectroscopy. *Proc. Natl. Acad. Sci. U.S.A.*, **108**, 16521 (2011).

[111] M. Lustig, D. Donoho, and J. M. Pauly. Sparse MRI: The application of compressed sensing for rapid MR imaging. *Magnet. Reson. Med.*, **58**, 1182 (2007).

[112] J. Ma. Compressed Sensing for Surface Characterization and Metrology. *IEEE Trans. Instrum. Meas.*, **59**, 1600 (2010).

[113] S. Mallat. *A Wavelet Tour of Signal Processing: The Sparse Way*. Academic Press, 3rd edition, 2008.

[114] T. Mančal, A. V. Pisliakov, and G. R. Fleming. Two-dimensional optical three-pulse photon echo spectroscopy. I. Nonperturbative approach to the calculation of spectra. *J. Chem. Phys.*, **124**, 234504 (2006).

[115] T. Markovich, S. M. Blau, J. Parkhill, C. Kreisbeck, J. N. Sanders, X. Andrade, and A. Aspuru-Guzik. More accurate and efficient bath spectral densities from super-resolution. arXiv:1307.4407 (2013).

[116] T. Markovich, S. M. Blau, J. N. Sanders, and A. Aspuru-Guzik. Benchmarking Compressed Sensing, Super-Resolution, and Filter Diagonalization. *Int. J. Quant. Chem.* (2016). In Press. arXiv:1502.06579.

[117] M. A. L. Marques, N. T. Maitra, F. M. S. Nogueira, E. K. U. Gross, and A. Rubio, editors. *Fundamentals of Time-Dependent Density Functional Theory*. Springer, 2012.

[118] D. Marx and J. Hutter. *Ab Initio Molecular Dynamics: Basic Theory and Advanced Methods*. Cambridge University Press, 2009.

[119] J. R. McClean and A. Aspuru-Guzik. Compact wavefunctions from compressed imaginary time evolution. *RSC Adv.*, **5**, 102277 (2015).

[120] C. W. McCutchen. Superresolution in microscopy and the Abbe resolution limit. *J. Opt. Soc. Am.*, **57**, 1190 (1967).

[121] B. Miehlich, A. Savin, H. Stoll, and H. Preuss. Results obtained with the correlation energy density functionals of Becke and Lee, Yang and Parr. *Chem. Phys. Lett.*, **157**, 200 (1989).

[122] M. Mishali and Y. C. Eldar. From Theory to Practice: Sub-Nyquist Sampling of Sparse Wideband Analog Signals. *IEEE J. Sel. Topics Signal Process.*, **4**, 375 (2010).

[123] M. Mobli, M. W. Maciejewski, A. D. Schuyler, A. S. Stern, and J. C. Hoch. Sparse Sampling Methods in Multidimensional NMR. *Phys. Chem. Chem. Phys.*, **14**, 10835 (2012).

[124] M. Mohseni, P. Rebentrost, S. Lloyd, and A. Aspuru-Guzik. Environment-assisted quantum walks in photosynthetic energy transfer. *J. Chem. Phys.*, **129**, 174106 (2008).

[125] J. Moix, J. Wu, P. Huo, D. Coker, and J. Cao. Efficient Energy Transfer in Light-Harvesting Systems, III: The Influence of the Eighth Bacteriochlorophyll on the Dynamics and Efficiency in FMO. *J. Phys. Chem. Lett.*, **2**, 3045 (2011).

[126] C. Møller and M. S. Plesset. Note on an Approximation Treatment for Many-Electron Systems. *Phys. Rev.*, **46**, 618 (1934).

[127] P. Mukherjee, I. Kass, I. T. Arkin, and M. T. Zanni. Picosecond dynamics of a membrane protein revealed by 2D IR. *Proc. Natl. Acad. Sci. U.S.A.*, **103**, 3528 (2006).

[128] L. J. Nelson, G. L. W. Hart, F. Zhou, and V. Ozolins. Compressive sensing as a paradigm for building physics models. *Phys. Rev. B*, **87**, 035125 (2013).

[129] C. Ochsenfeld and M. Head-Gordon. A reformulation of the coupled-perturbed self-consistent field equations entirely within a local atomic orbital density matrix-based scheme. *Chem. Phys. Lett.*, **270**, 399 (1997).

[130] A. Oka and L. Lampe. A compressed sensing receiver for bursty communication with UWB Impulse Radio. *Proc. IEEE Int. Conf. Ultra-WideBand*, **9**, 279 (2009).

[131] C. Olbrich, T. L. C. Jansen, J. Liebers, M. Aghtar, J. Strümpfer, K. Schulten, J. Knoester, and U. Kleinekathöfer. From Atomistic Modeling to Excitation Transfer and Two-Dimensional Spectra of the FMO Light-Harvesting Complex. *J. Phys. Chem. B*, **115**, 8609 (2011).

[132] C. Olbrich and U. Kleinekathöfer. Time-Dependent Atomistic View on the Electronic Relaxation in Light-Harvesting System II. *J. Phys. Chem. B*, **114**, 12427 (2010).

[133] P. Ordejón, D. A. Drabold, R. M. Martin, and S. Itoh. Linear Scaling Method for Phonon Calculations from Electronic Structure. *Phys. Rev. Lett.*, **75**, 1324 (1995).

[134] Y. Osamura, Y. Yamaguchi, P. Saxe, D. J. Fox, M. A. Vincent, and H. F. Schaefer. Analytic Second Derivative Techniques for Self-Consistent-Field Wave Functions. A New Approach to the Solution of the Coupled Perturbed Hartree-Fock Equations. *J. Mol. Struct.-Theochem.*, **103**, 183 (1983).

[135] L. A. Pachón and P. Brumer. Computational methodologies and physical insights into electronic energy transfer in photosynthetic light-harvesting complexes. *Phys. Chem. Chem. Phys.*, **14**, 10094 (2012).

[136] A. J. Patti, M. I. Sezan, and A. Murat Tekalp. Superresolution Video Reconstruction with Arbitrary Sampling Lattices and Nonzero Aperture Time. *IEEE Trans. Image Process.*, **6**, 1064 (1997).

[137] J. P. Perdew, K. Burke, and M. Ernzerhof. Generalized Gradient Approximation Made Simple. *Phys. Rev. Lett.*, **77**, 3865 (1996).

[138] A. Perdomo-Ortiz, J. R. Widom, G. A. Lott, A. Aspuru-Guzik, and A. H. Marcus. Conformation and Electronic Population Transfer in Membrane Supported Self-Assembled Porphyrin Dimers by 2D Fluorescence Spectroscopy. *J. Phys. Chem. B*, **116**, 10757 (2012).

[139] A. Pereverzev and E. R. Bittner. Time-convolutionless master equation for mesoscopic electron-phonon systems. *J. Chem. Phys.*, **125**, 104906 (2006).

[140] M. B. Plenio and S. F. Huelga. Dephasing-assisted transport: quantum networks and biomolecules. *New J. Phys.*, **10**, 113019 (2008).

[141] J. A. Pople, R. Krishnan, H. B. Schlegel, and J. S. Binkley. Derivative Studies in Hartree-Fock and Møller-Plesset Theories. *Int. J. Quantum Chem. Symp.*, **16**, 225 (1979).

[142] P. Pulay. Ab initio calculation of force constants and equilibrium geometries in polyatomic molecules. I. Theory. *Mol. Phys.*, **17**, 197 (1969).

[143] G. D. Purvis and R. J. Bartlett. A full coupled-cluster singles and doubles model: The inclusion of disconnected triples. *J. Chem. Phys.*, **76**, 1910 (1982).

[144] K. G. Puschmann and F. Kneer. On super-resolution in astronomical imaging. *Astron. Astrophys.*, **436**, 373 (2005).

[145] C. Qiu, W. Lu, and N. Vaswani. Real-time dynamic MR image reconstruction using Kalman Filtered Compressed Sensing. In *Proc. IEEE Int. Conf. Acoust. Speech Signal Process.*, volume 34, pages 393–396, 2009.

[146] D. C. Rapaport. *The Art of Molecular Dynamics Simulation*. Cambridge University Press, 1995.

[147] P. Rebentrost and A. Aspuru-Guzik. Communication: Exciton–phonon information flow in the energy transfer process of photosynthetic complexes. *J. Chem. Phys.*, **134**, 101103 (2011).

[148] P. Rebentrost, M. Mohseni, I. Kassal, S. Lloyd, and A. Aspuru-Guzik. Environment-assisted quantum transport. *New J. Phys.*, **11**, 033003 (2009).

[149] M. Reiher and J. Neugebauer. A mode-selective quantum chemical method for tracking molecular vibrations applied to functionalized carbon nanotubes. *J. Chem. Phys.*, **118**, 1634 (2003).

[150] G. Ritschel, J. Roden, W. T. Strunz, A. Aspuru-Guzik, and A. Eisfeld. Absence of Quantum Oscillations and Dependence on Site Energies in Electronic Excitation Transfer in the Fenna–Matthews–Olson Trimer. *J. Phys. Chem. Lett.*, **2**, 2912 (2011).

[151] J. Roden, A. Eisfeld, W. Wolff, and W. Strunz. Influence of Complex Exciton-Phonon Coupling on Optical Absorption and Energy Transfer of Quantum Aggregates. *Phys. Rev. Lett.*, **103**, 058301 (2009).

[152] R. Rubinstein, A. M. Bruckstein, and M. Elad. Dictionaries for Sparse Representation Modeling. *Proc. IEEE*, **98**, 1045 (2010).

[153] E. Runge and E. K. U. Gross. Density-Functional Theory for Time-Dependent Systems. *Phys. Rev. Lett.*, **52**, 997 (1984).

[154] J. N. Sanders, X. Andrade, and A. Aspuru-Guzik. Compressed Sensing for the Fast Computation of Matrices: Application to Molecular Vibrations. *ACS Cent. Sci.*, **1**, 2697 (2015).

[155] J. N. Sanders, S. K. Saikin, S. Mostame, X. Andrade, J. R. Widom, A. H. Marcus, and A. Aspuru-Guzik. Compressed Sensing for Multidimensional Spectroscopy Experiments. *J. Phys. Chem. Lett.*, **3**, 2697 (2012).

[156] M. Sarovar, A. Ishizaki, G. R. Fleming, and K. B. Whaley. Quantum entanglement in photosynthetic light-harvesting complexes. *Nature Phys.*, **6**, 462 (2010).

[157] H. Schaeffer, R. Caflisch, C. D. Hauck, and S. Osher. Sparse dynamics for partial differential equations. *Proc. Natl. Acad. Sci. U.S.A.*, **110**, 6634 (2013).

[158] H. L. Sellers, V. J. Klimkowski, and L. Schäfer. Normal Coordinate Ab Initio Force Relaxation. *Chem. Phys. Lett.*, **58**, 541 (1978).

[159] A. Shabani, R. L. Kosut, M. Mohseni, H. Rabitz, M. A. Broome, M. P. Almeida, A. Fedrizzi, and A. G. White. Efficient Measurement of Quantum Dynamics via Compressive Sensing. *Phys. Rev. Lett.*, **106**, 100401 (2011).

[160] A. Shabani and D. A. Lidar. Completely positive post-Markovian master equation via a measurement approach. *Phys. Rev. A*, **71**, 020101 (2005).

[161] Y. Shao, Z. Gan, E. Epifanovsky, A. T. B. Gilbert, M. Wormit, J. Kussmann, A. W. Lange, A. Behn, J. Deng, X. Feng, D. Ghosh, M. Goldey, P. R. Horn, L. D. Jacobson, I. Kaliman, R. Z. Khaliullin, T. Kuś, A. Landau, J. Liu, E. I. Proynov, Y. M. Rhee, R. M. Richard, M. A. Rohrdanz, R. P. Steele, E. J. Sundstrom, H. L. Woodcock, P. M. Zimmerman, D. Zuev, B. Albrecht, E. Alguire, B. Austin, G. J. O. Beran, Y. A. Bernard, E. Berquist, K. Brandhorst, K. B. Bravaya, S. T. Brown, D. Casanova, C.-M. Chang, Y. Chen, S. H. Chien, K. D. Closser, D. L. Crittenden, M. Diedenhofen, R. A. DiStasio, H. Do, A. D. Dutoi, R. G. Edgar, S. Fatehi, L. Fusti-Molnar, A. Ghysels, A. Golubeva-Zadorozhnaya, J. Gomes, M. W. D. Hanson-Heine, P. H. P. Harbach, A. W. Hauser, E. G. Hohenstein, Z. C. Holden, T.-C. Jagau, H. Ji, B. Kaduk, K. Khistyaev, J. Kim, J. Kim, R. A. King, P. Klunzinger, D. Kosenkov, T. Kowalczyk, C. M. Krauter, K. U. Lao, A. Laurent, K. V. Lawler, S. V. Levchenko, C. Y. Lin, F. Liu, E. Livshits, R. C. Lochan, A. Luenser, P. Manohar, S. F. Manzer, S.-P. Mao, N. Mardirossian, A. V. Marenich, S. A. Maurer, N. J. Mayhall, E. Neuscamman, C. M. Oana, R. Olivares-Amaya, D. P. O'Neill, J. A. Parkhill, T. M. Perrine, R. Peverati, A. Prociuk, D. R. Rehn, E. Rosta, N. J. Russ, S. M. Sharada, S. Sharma, D. W. Small, A. Sodt, T. Stein, D. Stck, Y.-C. Su, A. J. W. Thom, T. Tsuchimochi, V. Vanovschi, L. Vogt, O. Vydrov, T. Wang, M. A. Watson, J. Wenzel, A. White, C. F. Williams, J. Yang, S. Yeganeh, S. R. Yost, Z.-Q. You, I. Y. Zhang, X. Zhang, Y. Zhao, B. R. Brooks, G. K. L. Chan, D. M. Chipman, C. J. Cramer, W. A. Goddard, M. S. Gordon, W. J. Hehre, A. Klamt, H. F. Schaefer, M. W. Schmidt, C. D. Sherrill, D. G. Truhlar, A. Warshel, X. Xu, A. Aspuru-Guzik, R. Baer, A. T. Bell,

N. A. Besley, J.-D. Chai, A. Dreuw, B. D. Dunietz, T. R. Furlani, S. R. Gwaltney, C.-P. Hsu, Y. Jung, J. Kong, D. S. Lambrecht, W. Liang, C. Ochsenfeld, V. A. Rassolov, L. V. Slipchenko, J. E. Subotnik, T. V. Voorhis, J. M. Herbert, A. I. Krylov, P. M. W. Gill, and M. Head-Gordon. Advances in molecular quantum chemistry contained in the Q-Chem 4 program package. *Mol. Phys.*, **113**, 184 (2015).

[162] S. M. Sharada, A. T. Bell, and M. Head-Gordon. A finite difference Davidson procedure to sidestep full ab initio hessian calculation: Application to characterization of stationary points and transition state searches. *J. Chem. Phys.*, **140**, 164115 (2014).

[163] N. Shenvi, J. R. Schmidt, S. T. Edwards, and J. C. Tully. Efficient discretization of the continuum through complex contour deformation. *Phys. Rev. A*, **78**, 022502 (2008).

[164] S. Shim, P. Rebentrost, S. Valleau, and A. Aspuru-Guzik. Atomistic Study of the Long-Lived Quantum Coherences in the Fenna-Matthews-Olson Complex. *Biophys. J.*, **102**, 649 (2012).

[165] Y. Shrot and L. Frydman. Compressed sensing and the reconstruction of ultrafast 2D NMR data: Principles and biomolecular applications. *J. Magn. Reson.*, **209**, 352 (2011).

[166] N. Singh and P. Brumer. Electronic energy transfer in model photosynthetic systems: Markovian vs. non-Markovian dynamics. *Faraday Discuss.*, **153**, 41 (2011).

[167] N. Skochdopole and D. A. Mazziotti. Functional Subsystems and Quantum Redundancy in Photosynthetic Light Harvesting. *J. Phys. Chem. Lett.*, **2**, 2989 (2011).

[168] A. Smirne and B. Vacchini. Nakajima-Zwanzig versus time-convolutionless master equation for the non-Markovian dynamics of a two-level system. *Phys. Rev. A*, **82**, 022110 (2010).

[169] B. Song, N. Xi, H. Chen, K. W. C. Lai, and L. Chen. Carbon Nanotube-Based Infrared Camera Using Compressive Sensing. In N. Xi and K. W. C. Lai, editors, *Nano-Optoelectronic Sensors and Devices: Nanophotonics from Design to Manufacturing*, pages 225–243. Elsevier, 2012.

[170] D. A. Steck. *Rubidium 87 D Line Data*. Theoretical Division (T-8), MS B285, Los Alamos National Laboratory, Los Alamos, NM, 2001. Available online at http://steck.us/alkalidata.

[171] A. S. Stern, D. L. Donoho, and J. C. Hoch. NMR data processing using iterative thresholding and minimum l_1-norm reconstruction. *J. Magn. Reson.*, **188**, 295 (2007).

[172] T. Strohmer. Measure What Should be Measured: Progress and Challenges in Compressive Sensing. *IEEE Signal Process. Lett.*, **19**, 887 (2012).

[173] A. Szabo and N. S. Ostlund. *Modern Quantum Chemistry: Introduction to Advanced Electronic Structure Theory*. Dover Publications, 2nd edition, 1996.

[174] T. Takada, M. Dupuis, and H. F. King. Molecular symmetry. III. Second derivatives of electronic energy with respect to nuclear coordinates. *J. Chem. Phys.*, **75**, 332 (1981).

[175] Y. Takimoto, F. D. Vila, and J. J. Rehr. Real-time time-dependent density functional theory approach for frequency-dependent nonlinear optical response in photonic molecules. *J. Chem. Phys.*, **127**, 154114 (2007).

[176] J. Tang and J. R. Norris. LPZ spectral analysis using linear prediction and the z-transform. *J. Chem. Phys.*, **84**, 5210 (1986).

[177] Y. Tanimura and R. Kubo. Time Evolution of a Quantum System in Contact with a Nearly Gaussian-Markoffian Noise Bath. *J. Phys. Soc. Jpn.*, **58**, 101 (1989).

[178] P. F. Tekavec, G. A. Lott, and A. H. Marcus. Fluorescence-detected two-dimensional electronic coherence spectroscopy by acousto-optic phase modulation. *J. Chem. Phys.*, **127**, 214307 (2007).

[179] P. Tian, D. Keusters, Y. Suzaki, and W. S. Warren. Femtosecond Phase-Coherent Two-Dimensional Spectroscopy. *Science*, **300**, 1553 (2003).

[180] C. Timm. Time-convolutionless master equation for quantum dots: Perturbative expansion to arbitrary order. *Phys. Rev. B*, **83**, 115416 (2011).

[181] M. Tuckerman. *Statistical Mechanics: Theory and Molecular Simulation.* Oxford University Press, 2010.

[182] T. Tuma, S. Rooney, and P. Hurley. On the Applicability of Compressive Sampling in Fine Grained Processor Performance Monitoring. *Proc. IEEE Int. Conf. Eng. Complex Comput. Syst.*, **14**, 210 (2009).

[183] S. Valleau, A. Eisfeld, and A. Aspuru-Guzik. On the alternatives for bath correlators and spectral densities from mixed quantum-classical simulations. *J. Chem. Phys.*, **137**, 224103 (2012).

[184] E. van den Berg and M. P. Friedlander. Probing the Pareto frontier for Basis Pursuit Solutions. *SIAM J. Sci. Comput.*, **31**, 890 (2009).

[185] D. Varsano, L. A. Espinosa-Leal, X. Andrade, M. A. L. Marques, R. di Felice, and A. Rubio. Towards a gauge invariant method for molecular chiroptical properties in TDDFT. *Phys. Chem. Chem. Phys.*, **11**, 4481 (2009).

[186] J. C. Vaughan, T. Hornung, K. W. Stone, and K. A. Nelson. Coherently Controlled Ultrafast Four-Wave Mixing Spectroscopy. *J. Phys. Chem. A*, **111**, 4873 (2007).

[187] S. M. Vlaming and R. J. Silbey. Correlated intermolecular coupling fluctuations in photosynthetic complexes. *J. Chem. Phys.*, **136**, 055102 (2012).

[188] S. I. E. Vulto, M. A. de Baat, R. J. W. Louwe, H. P. Permentier, T. Neef, M. Miller, H. van Amerongen, and T. J. Aartsma. Exciton Simulations of Optical Spectra of the FMO Complex from the Green Sulfur Bacterium Chlorobium tepidum at 6 K. *J. Phys. Chem. B*, **102**, 9577 (1998).

[189] A. E. Waters, A. C. Sankaranarayanan, and R. G. Baraniuk. SpaRCS: Recovering Low-Rank and Sparse Matrices from Compressive Measurements. *Adv. Neural Inf. Process. Syst.*, **24**, 1089 (2011).

[190] E. B. Wilson, J. C. Decius, and P. C. Cross. *Molecular Vibrations: The Theory of Infrared and Raman Vibrational Spectra*. Dover Publications, 1980.

[191] J. Wright, A. Ganesh, K. Min, and Y. Ma. Compressive principal component pursuit. *Inf. Inference*, **2**, 32 (2013).

[192] J. Wu, F. Liu, Y. Shen, J. Cao, and R. J. Silbey. Efficient energy transfer in light-harvesting systems, I: optimal temperature, reorganization energy and spatial-temporal correlations. *New J. Phys.*, **12**, 105012 (2010).

[193] D. Xu, Y. Huang, and J. U. Kang. Real-time compressive sensing spectral domain optical coherence tomography. *Opt. Lett.*, **39**, 76 (2013).

[194] K. Yabana and G. F. Bertsch. Time-dependent local-density approximation in real time. *Phys. Rev. B*, **54**, 4484 (1996).

[195] K. Yabana and G. F. Bertsch. Application of the time-dependent local density approximation to optical activity. *Phys. Rev. A*, **60**, 1271 (1999).

[196] K. Yabana, T. Nakatsukasa, J.-I. Iwata, and G. F. Bertsch. Real-time, real-space implementation of the linear response time-dependent density-functional theory. *Phys. Status Solidi B*, **243**, 1121 (2006).

[197] Y. Yamaguchi, Y. Osamura, J. D. Goddard, and H. F. Schaefer. *A New Dimension to Quantum Chemistry: Analytic Derivative Methods in Ab Initio Molecular Electronic Structure Theory*. Oxford University Press, 1994.

[198] J. Yuen-Zhou and A. Aspuru-Guzik. Quantum process tomography of excitonic dimers from two-dimensional electronic spectroscopy. I. General theory and application to homodimers. *J. Chem. Phys.*, **134**, 134505 (2011).

[199] J. Yuen-Zhou, J. J. Krich, and A. Aspuru-Guzik. A witness for coherent electronic vs vibronic-only oscillations in ultrafast spectroscopy. *J. Chem. Phys.*, **136**, 234501 (2012).

[200] J. Yuen-Zhou, J. J. Krich, M. Mohseni, and A. Aspuru-Guzik. Quantum state and process tomography of energy transfer Systems via ultrafast spectroscopy. *Proc. Natl. Acad. Sci. U.S.A.*, **108**, 17615 (2011).

[201] Z. Zhang, L. Piatkowski, H. J. Bakker, and M. Bonn. Ultrafast vibrational energy transfer at the water/air interface revealed by two-dimensional surface vibrational spectroscopy. *Nat. Chem.*, **3**, 888 (2011).

[202] Y. Zhao and D. G. Truhlar. The M06 suite of density functionals for main group thermochemistry, thermochemical kinetics, noncovalent interactions, excited states, and transition elements: two new functionals and systematic testing of four M06-class functionals and 12 other functionals. *Theor. Chem. Account.*, **120**, 215 (2006).

[203] Z. Zhou, X. Li, J. Wright, E. Candès, and Y. Ma. Stable Principal Component Pursuit. *IEEE Int. Symp. Inf. Theory Proc.*, pages 1518–1522 (2010).

[204] J. Zhu, S. Kais, P. Rebentrost, and A. Aspuru-Guzik. Modified Scaled Hierarchical Equation of Motion Approach for the Study of Quantum Coherence in Photosynthetic Complexes. *J. Phys. Chem. B*, **115**, 1531 (2011).

[205] L. Zhu, W. Zhang, D. Elnatan, and B. Huang. Faster STORM using compressed sensing. *Nat. Methods*, **9**, 721 (2012).